

John E. Bradshaw

# Plant Breeding: Past, Present and Future



Springer

# Plant Breeding: Past, Present and Future



John E. Bradshaw

# Plant Breeding: Past, Present and Future



Springer

John E. Bradshaw  
Edinburgh, UK

ISBN 978-3-319-23284-3      ISBN 978-3-319-23285-0 (eBook)  
DOI 10.1007/978-3-319-23285-0

Library of Congress Control Number: 2015951092

© Springer International Publishing Switzerland 2016

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

This Springer imprint is published by Springer Nature  
The registered company is Springer International Publishing AG Switzerland

# Preface

Plant breeders starting work in 2015 will be nearing the end of their careers in 2050. Both years are of significance. The United Nations set eight ambitious Millennium Development Goals to be achieved by 2015, from a 1990 baseline. One was halving extreme poverty rates and included halving the proportion of people who suffer from hunger. The commitment to halve the percentage of hungry people by 2015 has almost been met at the global level and has been achieved by a total of 72 developing countries out of 129, or more than half the countries monitored according to the 2015 report from FAO, IFAD and WFP (2015). The report estimates that about 795 million people (780 million in the developing regions) are undernourished globally, which is a reduction of 167 million over the last decade and 216 million less than in 1990–1992. For the developing regions as a whole, the share of undernourished people in the total population has decreased from 23.3 % in 1990–1992 to 12.9 %. The report concluded that economic growth was a key factor for reducing undernourishment, but had to be inclusive and provide opportunities for improving the livelihoods of the poor. Enhancing the productivity and incomes of smallholder family farmers remains a key to progress. Encouraging as this is, a tremendous amount of work remains to be done to eliminate hunger, and the situation could worsen between 2015 and 2050.

The United Nations predicts that the global human population will continue to rise from the 7.0 billion reached in 2011 to 9.0 billion by 2050, when 70 % of the population will live in urban environments. As a consequence, world food production will need to increase between 70 and 100 % in just 40 years. Once again there are fears of population growth overtaking food supplies, as famously predicted by Thomas Robert Malthus in 1798 in *An Essay on the Principle of Population* (Malthus, 1798), when the population was only around 1.0 billion. Increasing food production to feed 9.0 billion people will be made even more challenging by four factors. Firstly, urbanisation and increasing wealth in developing countries (particularly China) is leading to a shift in diets to meat and dairy products, and these require the use of more farmland than an equivalent plant-based diet. Secondly, first-generation biofuels are using crops and cropland to produce energy

rather than food so that even more food will need to be produced from the remaining land. Thirdly, the land area used for agriculture may remain static or even decrease as a result of degradation and climate change, despite more land being theoretically available, unless crops can be bred which tolerate associated abiotic stresses. Fourthly, it is unlikely that steps can be taken to mitigate all of the climate change predicted to occur by 2050, and beyond, and hence adaptation of farming systems and crop production will be required to reduce predicted negative effects on yields that will occur without crop adaptation. These impacts have been analyzed and summarised in the Intergovernmental Panel on Climate Change Report (AR5) published on 31 March 2014 (Climate Change: Impacts, Adaptation, and Vulnerability) and included in its Synthesis Report published on 31 October 2014. Substantial progress will be required in bridging the yield gap between what is currently achieved per unit of land and what should be possible in the future, given the most appropriate farming methods and best storage and transportation of food. Suitably adapted cultivars will also be needed, including adaptation to climate change (e.g. cultivars adapted to different growing seasons and to heat stress).

Breeding suitably adapted cultivars is the challenge for a new generation of plant breeders. They will need to decide what germplasm and which breeding methods to use and the types of cultivar to produce. They will need to appreciate what germplasm has already been utilised and how to find further desirable genetic variants in cultivated species and their wild relatives. They will also need to appreciate which breeding methods have been successful in the past and which ones are still likely to be successful today. In addition, they will need to consider the new opportunities made possible by technological advances in the manipulation of DNA, the chemical basis of heredity.

This book aims to help plant breeders in these endeavours by reviewing past achievements, currently successful practices and emerging methods and techniques. Theoretical considerations are presented when thought helpful and trying to strike the right balance on being as simple as possible but as complex as necessary. On a number of occasions, I suggest that practical breeders may get more out of computer simulations than complex theory. The book assumes an understanding of botany, genetics and statistics and does not attempt to teach these subjects, although reminders are given of facts relevant to plant breeding. However, as I have discovered from teaching undergraduate and postgraduate courses, it is impossible to discuss plant breeding in a meaningful way without presenting some genetic theory, particularly in population and quantitative genetics. Furthermore, the reader may need to consult a *Dictionary of Plant Sciences*, such as the one by Allaby (2012), for the meaning of some terms.

The book is divided into four parts. Part I is an historical introduction, from domestication to landraces and scientific breeding in the twentieth century, finishing with a discussion of future goals. Part II deals with the origin of genetic variation by mutation and recombination of DNA, the recognition and selection of genetic variants that affect qualitative and quantitative traits in a desired way and the implications of genotype  $\times$  environment interactions for selection. This part concludes with a consideration of genome evolution and polyploidy, including

examples of crop species that are polyploids and any implications for breeding. Part III starts by explaining how the mating system of a crop species determines the genetic structure of its landraces and hence the types of high yielding cultivars that have been selected and bred from landraces. Then current practice is reviewed for breeding the four main types of cultivar: synthetic (including open-pollinated), clonal, hybrid and inbred line (including mixtures). Finally the genetic basis of heterosis is examined in the context of choosing between breeding inbred line and hybrid cultivars. Opportunities are taken in this part to introduce important aspects of the nature of plant breeding: population improvement, multistage selection, multitrait selection, inbreeding and crossbreeding, hybridisation strategy and size of programme. Part IV considers the three complementary options for future progress: the use of sexual reproduction in further conventional breeding, base broadening and introgression; mutation breeding; and genetically modified crops. It then concludes with strategies for achieving durable resistance to pests and diseases, something that has so often appeared to be an historical fact or just hope for the future.

In writing the book I have drawn on my experiences over a 34-year period as a plant breeder and geneticist at the former Scottish Plant Breeding Station in Edinburgh and the former Scottish Crop Research Institute in Dundee. Throughout my career I have benefited from discussions with colleagues and the wider plant breeding community, particularly members of EUCARPIA, the European Association for Research on Plant Breeding. I have also benefited from participation in teaching, working groups and editorial work. Recently I edited Volume 7 (*Root and Tuber Crops*) in Springer's *Handbook of Plant Breeding*. In writing this book I have once again received much help and encouragement from Springer, this time from Ejaz Ahmad, Susan Westendorf and Kenneth Teng in particular. The 'handbook' provides detailed information on a wide range of economically important crops and should be of great value to breeders of those crops. In addition, I feel that those new to plant breeding might appreciate a more general review of the subject, its past successes, current practices and future possibilities. I hope that this book goes some way to meeting this objective.

Edinburgh, UK

June 2015

John E. Bradshaw



# Contents

<b>Prologue . . . . .</b>	<b>xxv</b>
---------------------------	------------

## Part I Historical Introduction

<b>1 Domestication, Dispersion, Selection and Hybridization of Cultivated Plants . . . . .</b>	<b>3</b>
Introduction . . . . .	3
Involvement of Humans . . . . .	3
Transition to Farming in the Near East (Southwest Asia) . . . . .	4
Bringing Land into Cultivation . . . . .	5
Domestication . . . . .	8
Centres of Diversity and Origin . . . . .	8
Features of Domestication . . . . .	10
Continuing Debates About Domestication . . . . .	11
Plants and People . . . . .	14
Dispersion . . . . .	15
Initial Spread of Farming . . . . .	15
Exchanges Between New and Old Worlds . . . . .	16
Colonialism and Tropical Plantation Crops . . . . .	19
Adaptation . . . . .	21
Selection During Domestication . . . . .	21
Cereal Domestication . . . . .	21
Potato Domestication . . . . .	22
Visual Selection . . . . .	22
Potatoes, Tomatoes and Carrots . . . . .	22
Selection in New Environments . . . . .	24
Selection for Use in New Farming and Production Systems . . . . .	25
Selection for Altered Chemical Composition . . . . .	26
Plant Breeders and Commodity Traders . . . . .	28
Sexual Reproduction in Flowering Plants . . . . .	29

Classification and Naming of Plants . . . . .	31
Species Concepts . . . . .	32
Deliberate Hybridization in Plant Breeding . . . . .	33
Mechanics of Artificial Hybridization . . . . .	33
Mendel and the Laws of Inheritance . . . . .	34
Darwin and the Evolutionary Significance of the Mating System . . . . .	34
Purity and Seed Quality of New Cultivars . . . . .	35
Cellular Basis of Life . . . . .	36
Cell Division . . . . .	38
<b>2 Scientific Breeding in the Twentieth Century and Future Goals . . . . .</b>	<b>39</b>
Genetics and Cytology (Cytogenetics): The Foundations of Scientific Breeding . . . . .	39
UPOV . . . . .	39
Continuous Variation and QTLs . . . . .	40
Polyploids . . . . .	41
Mutagenesis . . . . .	42
DNA and Plant Genomes . . . . .	43
Genetic Transformation . . . . .	45
Achievements in the Twentieth Century . . . . .	45
Hybrid Maize . . . . .	46
The Green Revolution . . . . .	47
Wheat in Mexico and India . . . . .	48
Wheat in China . . . . .	48
Rice in the Philippines and Asia . . . . .	49
Economics and Politics . . . . .	50
International Research Centres . . . . .	50
Increased Potato Production in China and India . . . . .	51
Genetically Modified Crops . . . . .	51
Reconnecting Breeders with Farmers and End Users . . . . .	53
Future Goals of Plant Breeding: Quantity and Quality of Food Crops . . . . .	53
World Food Production in 2005–2007 and Future Trends . . . . .	55
Feeding the World’s Population in 2050 . . . . .	56
Impact of Population Growth . . . . .	57
Factors Affecting Food Production . . . . .	58
Sustainable Food Production . . . . .	61
Available Inputs and Their Utilization by Plants . . . . .	62
Photosynthesis, Water, Temperature and Climate Change . . . . .	62
Climate Change . . . . .	64
Water Use . . . . .	65
Temperature . . . . .	67
Fertilizers and Nitrogen Fixation . . . . .	68
Reducing Yield Losses from Weeds, Pests and Diseases . . . . .	69
Goals, Selection Criteria and Success . . . . .	71

**Part II Reproduction, Heredity and Variation**

<b>3 DNA and the Origin of Variation . . . . .</b>	<b>75</b>
Introduction . . . . .	75
The Chemical and Physical Basis of Heredity and Variation . . . . .	75
Next-Generation Sequencing of DNA . . . . .	77
Life Cycle of Multicellular Plants . . . . .	77
Life Cycle of Angiosperms . . . . .	79
Pollen Grain . . . . .	80
Embryo Sac . . . . .	80
Fertilization and Development . . . . .	80
Cell Lineages and Chimeras . . . . .	81
Seeds and Fruits . . . . .	82
Six-Rowed and Two-Rowed Barley . . . . .	83
Wood . . . . .	83
Cytological Techniques for Studying Chromosomes . . . . .	84
FISH and GISH . . . . .	84
Mitosis . . . . .	85
Meiosis . . . . .	87
Key Interpretations of Meiosis for Heredity . . . . .	88
Molecular Basis of Genetic Recombination . . . . .	89
Changes in Chromosome Number: Errors in Distribution . . . . .	92
Importance of Meiosis in Plant Breeding . . . . .	93
DNA Damage . . . . .	93
DNA Repair . . . . .	94
Photoreactivation . . . . .	94
Excision Repair . . . . .	95
Double Strand Break Repair by Non Homologous End Joining . . . . .	95
Double Strand Break Repair by Homologous Recombination . . . . .	96
Mutations, the Source of Genetic Variation . . . . .	97
Base-Pair Substitutions . . . . .	97
RFLP Markers . . . . .	98
Markers Produced by PCR (Variation in DNA Sequence) . . . . .	98
Micro-Satellite Markers (Variation in DNA Repeats) . . . . .	100
Diversity Array Technology (DArT) Markers . . . . .	100
Deletions, Inversions and Translocations . . . . .	101
Inversions in the Monkeyflower . . . . .	101
Chromosomal Rearrangements in Wheat . . . . .	103
Chromosomal Evolution in the Plant Family Solanaceae . . . . .	105
Transposable Elements . . . . .	106
<b>4 Mendelian Genetics and Linkage Maps . . . . .</b>	<b>109</b>
Introduction . . . . .	109
Segregation . . . . .	109

Independent Assortment . . . . .	110
The Binomial Distribution . . . . .	111
<i>Chi-Square (<math>X^2</math>) Goodness-of-Fit Statistic</i> . . . . .	113
Terminology . . . . .	115
Partial Linkage and Linkage Maps . . . . .	116
Maize . . . . .	116
Ordered Tetrad Analysis . . . . .	118
Interference . . . . .	120
Estimation of Recombination Frequency . . . . .	121
Backcross . . . . .	121
Complications and the EM Algorithm . . . . .	122
Mapping Functions . . . . .	124
Test for Linkage and Membership of Linkage Group . . . . .	127
Ordering Loci in Linkage Group . . . . .	128
High Density SNP Maps . . . . .	130
Some Conclusions About High Density Maps . . . . .	131
<b>5 Gene Expression and Selection of Major Genes</b> . . . . .	133
Introduction . . . . .	133
Biochemical Pathways . . . . .	133
Epistasis and Modified Mendelian Ratios . . . . .	134
Cyanogenesis . . . . .	135
The Genetic Code . . . . .	135
Herbicide Tolerant Crops . . . . .	137
Control of Gene Expression . . . . .	139
Promoter . . . . .	139
Transcription Factors . . . . .	140
Mendel's Genes . . . . .	141
Domestication Genes . . . . .	142
Maize . . . . .	143
Rice . . . . .	143
Sorghum . . . . .	144
Wheat . . . . .	145
Barley . . . . .	146
Number of Domestication Genes . . . . .	147
Adaptation to New Environments . . . . .	148
Photoperiod: Wheat . . . . .	148
Photoperiod: Maize . . . . .	148
Maturity: Potato . . . . .	149
Vernalization: Wheat and Barley . . . . .	150
Selection for Colour and Shape . . . . .	152
Colour: Tuber Skin of Potato . . . . .	152
Colour: Red and White Grapes . . . . .	153
Shape: Tomatoes . . . . .	154

Green Revolution Genes . . . . .	155
Wheat . . . . .	155
Rice . . . . .	156
Genome-Wide Analysis of Gene Expression . . . . .	157
Microarrays . . . . .	157
Heritable Epigenetic Variation Among Maize Inbreds . . . . .	158
<b>6 Quantitative Genetics and Genomic Selection . . . . .</b>	<b>161</b>
Introduction . . . . .	161
Quantitative Traits . . . . .	163
Yield Trials . . . . .	166
The General Mixed Model and Estimation of Breeding Values . . . . .	169
Plot Interference . . . . .	174
High-Throughput Phenotyping . . . . .	176
Quantitative Trait Loci (QTLs) . . . . .	177
Backcross Population: Graphical Genotyping . . . . .	177
Backcross Population: QTL Analysis . . . . .	178
Interval Mapping . . . . .	179
The Problem of Two Linked QTLs . . . . .	181
Marker-Assisted Selection Versus Phenotypic Selection in Winter Wheat . . . . .	183
Permutation Tests . . . . .	184
Composite Interval Mapping . . . . .	185
Introgression Mapping in <i>Lolium/Festuca</i> System . . . . .	186
Chromosome Segment Substitution Lines (CSSLs) . . . . .	187
Peanut . . . . .	189
Rice . . . . .	190
Mapping of QTL by Whole Genome Resequencing of DNA from Two Bulked Populations . . . . .	191
Candidate Genes . . . . .	192
Expression Quantitative Trait Loci (eQTLs) . . . . .	193
Association Genetics and Genomic Selection . . . . .	193
A Simple Example . . . . .	194
Issues in Practice for GWAS and GS . . . . .	197
Linkage Disequilibrium . . . . .	198
Population Structure . . . . .	198
Kinship . . . . .	201
Population Structure and Kinship . . . . .	201
GS Methods . . . . .	202
Cross-Validation . . . . .	203
GS Results in Practice . . . . .	203
Genomic Prediction in Tropical Maize Breeding . . . . .	205
Implications for Breeders . . . . .	206

<b>7 Genotype x Environment Interactions and Selection Environments . . . . .</b>	207
Introduction . . . . .	207
Concept of Genotype × Environment Interactions ( $G \times E$ ) . . . . .	209
Detection and Analysis . . . . .	210
Yield Potential and Drought Resistance or Tolerance . . . . .	211
Salt Tolerance and Yield . . . . .	213
Regression Analysis . . . . .	214
Selecting Genotypes in Different Types of Environment . . . . .	216
Response Curves and Deviations from Regression . . . . .	216
Nitrogen Use Efficiency . . . . .	218
Phosphorus Use Efficiency . . . . .	220
Principal Component Analysis . . . . .	221
$G \times E$ as a Correlated Response to Selection . . . . .	227
Conventional Versus Organic Farming . . . . .	228
Sorghum in West Africa . . . . .	229
Days to Flowering and Photoperiod Sensitivity . . . . .	230
Maize . . . . .	230
Rice . . . . .	231
<b>8 Genome Evolution and Polyploidy . . . . .</b>	233
Introduction . . . . .	233
Molecular Evolution and Phylogenetic Trees . . . . .	233
A Markov Model of Base Substitutions . . . . .	234
The Jukes-Cantor Model . . . . .	235
Jukes-Cantor Phylogenetic Distance . . . . .	237
Phylogenetic Trees . . . . .	238
Chloroplast and Mitochondrial Genomes . . . . .	240
Classification of Polyploids . . . . .	241
Unreduced Gametes from Diploid Species . . . . .	242
Potatoes . . . . .	242
Unreduced Gametes and Sexual Polyploidization . . . . .	243
The Genetic Consequences of FDR and SDR in Diploids . . . . .	244
Tetrasomic Inheritance in Autotetraploids . . . . .	248
Meiosis in Autotetraploids . . . . .	248
Double Reduction in Autotetraploids . . . . .	249
Linkage Analysis with Chromosomal Segregation . . . . .	250
QTL Analysis in Autotetraploids . . . . .	253
Examples of Polyploid Species . . . . .	254
Banana and Citrus Species (Triploid) . . . . .	255
Banana . . . . .	255
Citrus Species . . . . .	257
Brassicas (Diploids and Allotetraploids) and Raphanobrassica . . . . .	258
<i>Arabidopsis–Brassica</i> Genome Evolution . . . . .	258
Raphanobrassica . . . . .	259

<i>Festolium</i> (Allotetraploid) . . . . .	260
Cereal Genome Evolution . . . . .	261
Wheat (Allohexaploid) . . . . .	262
Triticale: A New Crop . . . . .	263
Complex Polyploids: Sugarcane . . . . .	264
Inbred Lines of Maize and the Pan-Genome Concept . . . . .	266
Copy Number and Present/Absent Variants . . . . .	267
Pan-Genome Concept . . . . .	267
Maize Subgenomes . . . . .	268
Maize Pan-Transcriptome . . . . .	269
Conclusions for Plant Breeders . . . . .	269

### Part III Landraces to High Yielding Cultivars

<b>9 Genetic Structure of Landraces . . . . .</b>	273
Introduction . . . . .	273
Mixtures of Inbred Lines, Hybrids and Clones . . . . .	274
Sexual Reproduction with Predominantly Self-Pollination . . . . .	275
Sexual Reproduction with Predominantly Cross-Pollination . . . . .	276
Sexual Reproduction with a Mixture of Self- and Cross-Pollination . . . . .	277
Sexual Reproduction Followed by Vegetative Propagation . . . . .	278
Concluding Remarks . . . . .	280
Mixed Selfing and Random Mating: Single Locus . . . . .	281
Mixed Selfing and Random Mating: Selection . . . . .	284
Mixed Selfing and Random Mating: Two Loci . . . . .	285
Random Mating: Two Loci . . . . .	287
Linkage Disequilibrium Under Mixed Selfing and Random Mating . . . . .	289
Mixed Selfing and Random Mating: Tetrasomic Inheritance . . . . .	289
Concluding Remarks . . . . .	290
<b>10 Open-Pollinated and Synthetic Cultivars from Population     Improvement . . . . .</b>	291
Introduction . . . . .	291
Landraces to Open-Pollinated Cultivars . . . . .	291
Maize . . . . .	291
Perennial Forage Crops and Amenity Grasses . . . . .	293
Foundation Population . . . . .	293
Population Improvement . . . . .	294
Mass, Half-Sib Family, Full-Sib Family and S <sub>1</sub> Family Selection . . . . .	296
Theory Under Disomic Inheritance . . . . .	296
Response to Selection . . . . .	299
Theory Under Tetrasomic Inheritance . . . . .	301
Making Choices . . . . .	306

Population Improvement in Practice . . . . .	307
Maize . . . . .	307
Perennial Ryegrass . . . . .	308
Alfalfa . . . . .	309
Red Clover . . . . .	310
White Clover . . . . .	311
Mixed Selfing and Random Mating . . . . .	312
Linkage Disequilibrium . . . . .	312
Epistasis . . . . .	316
Immediate Versus Equilibrium Response to Selection . . . . .	318
The Importance of Epistasis in Practice . . . . .	319
Selection for More Than One Trait . . . . .	319
Selection Indices . . . . .	320
Optimum Index of Smith for Half-Sib Family	
Selection in Kale . . . . .	320
Long-Term Selection in Maize . . . . .	323
Illinois Long Term Selection Experiment . . . . .	323
Conclusions from Selection Experiment . . . . .	325
Theory of Selection Limit . . . . .	326
Loss of Desirable Alleles by Genetic Drift . . . . .	327
Theory . . . . .	327
Computer Simulation . . . . .	329
Cultivar Production: Open-Pollinated Cultivars . . . . .	330
Cultivar Production: Synthetic Cultivars . . . . .	331
Prediction of Performance of Synthetic Cultivars: A Simple	
Diploid Example . . . . .	332
Prediction of Performance of Synthetic Cultivars:	
Some More General Diploid Results . . . . .	333
Complications in Autotetraploids . . . . .	338
Faba Bean Synthetics . . . . .	340
<b>11 Clonal Cultivars from Multistage Multitrait Selection . . . . .</b>	<b>343</b>
Introduction . . . . .	343
Hybridization Strategy . . . . .	344
Choice of Hybridizations . . . . .	344
Discarding Inferior Progenies . . . . .	345
Number of Crosses and Progeny Sizes . . . . .	345
Intensity of Selection . . . . .	345
Theoretical Example . . . . .	346
Theory of Clonal Selection: Multistage . . . . .	348
Effect of Environmental Variation on Ability	
to Select Best Clone . . . . .	348
Maximizing the Gain from Selection . . . . .	350
Single-Stage and Two-Stage Selection . . . . .	351
Other Approaches . . . . .	353

Theory of Clonal Selection: Multitrait . . . . .	354
Practice of Multitrait Selection . . . . .	357
Uncorrelated Traits . . . . .	358
Root and Tuber Crops . . . . .	359
Potato Breeding . . . . .	360
Modern Potato Breeding . . . . .	360
Scientific Potato Breeding in Scotland . . . . .	360
Breeding Scheme . . . . .	362
Analysis of Breeding Scheme: Crossing Strategy . . . . .	363
Analysis of Breeding Scheme: Early Generation Selection . . . . .	364
Genetic Modification, Marker-Assisted Selection, Genomic Selection and Phenotypic Selection . . . . .	366
Seed Potato Production . . . . .	366
Cassava . . . . .	368
Parents, Crossing and Clonal Generations . . . . .	368
Possible Improvements to Cassava Breeding . . . . .	370
Production of Planting Material . . . . .	370
Sweet Potato . . . . .	371
Tropical Plantation Crops . . . . .	372
Sugarcane . . . . .	372
Parents, Crossing and Clonal Generations . . . . .	374
Olive Breeding . . . . .	375
Fruit Breeding . . . . .	376
Strawberry Breeding . . . . .	377
Raspberry Breeding . . . . .	378
Grape Breeding . . . . .	379
Apple Breeding . . . . .	382
Genomic Selection in Apple Breeding . . . . .	384
Clonal Cultivars Produced by Apomixis . . . . .	385
Conclusions . . . . .	386
<b>12 Hybrid Cultivars from Inbreeding and Crossbreeding . . . . .</b>	<b>387</b>
Introduction . . . . .	387
Inbreeding and Crossbreeding . . . . .	388
Inbreeding . . . . .	388
Crossbreeding . . . . .	391
Variation Among the Inbred Lines and Single Cross Hybrids . . . . .	392
F <sub>1</sub> Hybrid Maize Breeding . . . . .	394
Selecting Inbred Lines to Produce the Best Possible Single Cross Hybrids . . . . .	394
Scheme for Development of Inbred Lines and Hybrids . . . . .	395
Genetic Basis of Selecting Inbreds . . . . .	397
Inbred Line <i>per se</i> Versus Testcross Selection . . . . .	399
Hybrid Maize Seed Production . . . . .	401
Cytoplasmic Male Sterility . . . . .	402
Heterotic Groups . . . . .	404

Theory . . . . .	404
Predicting Heterosis . . . . .	406
Heterotic Groups in Maize . . . . .	409
Reciprocal Recurrent Selection . . . . .	411
Theory . . . . .	411
Practice . . . . .	411
Botanical Varieties of <i>Brassica oleracea</i> . . . . .	413
Kale and Vegetables of <i>Brassica oleracea</i> . . . . .	414
Kale . . . . .	414
Vegetables of <i>Brassica oleracea</i> . . . . .	415
Onions and Carrots . . . . .	417
Onions . . . . .	417
The Use of CMS-S in Hybrid Onion Breeding . . . . .	418
A Hybrid Onion Breeding Scheme . . . . .	418
Hybrid Onion Seed Production . . . . .	419
Carrots . . . . .	420
Rye . . . . .	421
Sugar Beet . . . . .	422
<b>13 Inbred Line Cultivars and Mixtures from Hybridization and Inbreeding . . . . .</b>	<b>425</b>
Introduction . . . . .	425
Pure-Line Selection . . . . .	425
Artificial Hybridization . . . . .	426
Genetic Variation Between and Within Crosses . . . . .	427
Theory . . . . .	427
Conclusions from Theory . . . . .	429
Experimental Results . . . . .	430
Initial Choice of Crosses . . . . .	431
Discarding Crosses at F <sub>1</sub> and F <sub>3</sub> . . . . .	433
Use of F <sub>3</sub> Generation to Predict Recombinant Inbred Lines at F <sub>∞</sub> Generation . . . . .	435
Number of Crosses and Population Size of Progenies . . . . .	439
Past and Current Practice . . . . .	440
Pedigree Inbreeding with Selection, Single Seed Descent, Doubled Haploids and Bulk Population Selection . . . . .	440
Pedigree Inbreeding with Selection . . . . .	440
Single Seed Descent . . . . .	442
Doubled Haploids . . . . .	443
Bulk Population Selection . . . . .	444
Best Method of Producing Inbred Lines . . . . .	445
Selection Theory . . . . .	446
Comparison of Doubled Haploids and Lines from Single Seed Descent . . . . .	446
Selection During Pedigree Inbreeding . . . . .	448

Response to Selection in the F <sub>2</sub> Generation . . . . .	449
Effect of Selection in the F <sub>2</sub> Generation on the Variance . . . . .	449
Outcome of Selection at F <sub>2</sub> on the F <sub>∞</sub> Generation . . . . .	451
Outcome of a Single Round of Selection . . . . .	451
Selection in Two or More Generations . . . . .	453
Correlated Responses to Selection . . . . .	454
Value of Computer Simulations . . . . .	454
Conclusions on Breeding Methods . . . . .	455
Cultivar (Variety) Mixtures . . . . .	456
Rice Example . . . . .	456
Mixture Design . . . . .	458
Breeding for Mixing Ability . . . . .	458
<b>14 Genetic Basis of Heterosis and Inbred Line</b>	
<b>Versus Hybrid Cultivars</b> . . . . .	459
Introduction . . . . .	459
The Genetic Basis of Heterosis . . . . .	460
A Simple Example . . . . .	460
Dispersion of Dominant or Partially Dominant Genes . . . . .	462
Average Dominance Ratio at Individual Loci . . . . .	462
Linkage and Pseudo-Overdominance . . . . .	466
Epistasis . . . . .	467
Recent Examples of Research on the Genetic Basis of Heterosis . . . . .	469
Maize . . . . .	470
Rice . . . . .	470
Tomato . . . . .	472
Examples of Hybrid Breeding . . . . .	472
Grain Sorghum . . . . .	473
Rice . . . . .	474
Tomato . . . . .	475
<i>Brassica napus</i> Rapeseed . . . . .	476
Sunflower . . . . .	478
<b>Part IV Improving High Yielding Cultivars</b>	
<b>15 Use of Sexual Reproduction in Base Broadening and Introgression</b> . . . . .	483
Introduction . . . . .	483
Genetic Resources . . . . .	484
International Treaty on Plant Genetic Resources . . . . .	486
Global Crop Diversity Trust . . . . .	487
Svalbard Global Seed Vault . . . . .	488
Conservation . . . . .	488
Conservation <i>In Situ</i> . . . . .	488
Conservation <i>Ex Situ</i> . . . . .	490

Genebank Standards and Information . . . . .	491
Size of Sample When Collecting Germplasm . . . . .	492
Core Collections . . . . .	493
Cultivated Potatoes in Latin America: A Core Collection . . . . .	494
Diversity Fixed Foundation Sets . . . . .	495
Utilization of Genetic Resources . . . . .	495
Past Utilization of Genetic Resources . . . . .	496
Gene Pools . . . . .	498
Protoplast Fusion in <i>Nicotiana</i> . . . . .	499
Taxonomic Classification and Crossability in the Potato . . . . .	500
Taxonomy . . . . .	500
Crossability . . . . .	501
Introgression in Potato . . . . .	503
Introgression in Sugar Beet . . . . .	505
Introgression in Onion . . . . .	506
Introgression in Lettuce and Tomato . . . . .	507
Lettuce . . . . .	507
Tomato . . . . .	508
New Rice for Africa and Linkage Drag in Rice . . . . .	509
Molecular Marker-Assisted Introgression . . . . .	510
Chromosome Manipulation (Engineering) in Wheat . . . . .	516
Management of Homoeologous Pairing . . . . .	517
Translocation Breeding . . . . .	518
Base Broadening by Incorporation . . . . .	520
Base Broadening of Potatoes . . . . .	520
Base Broadening of Potatoes in Europe and North America . . . . .	520
CIP Populations B1 to B3 . . . . .	522
Germplasm Enhancement of Maize . . . . .	522
Germplasm Enhancement of Maize Project . . . . .	524
Historical Genomics of North American Maize . . . . .	525
Participatory Plant Breeding and Diversity . . . . .	526
<b>16 Mutation Breeding . . . . .</b>	<b>529</b>
Introduction . . . . .	529
Mutagenesis of Organelle Genomes . . . . .	531
Design of Mutation Experiments and Breeding Programmes . . . . .	532
Choice of Mutagen and Plant Material . . . . .	532
Doses and Dose Rates . . . . .	533
Dissociation of Chimeric Structure of M <sub>1</sub> Plants . . . . .	533
Number of Initial Cells . . . . .	534
Population Sizes in Mutation Programme with Vegetatively Propagated Crops . . . . .	535
Population Sizes in Mutation Programme with Self-Pollinated Crops . . . . .	536
Estimating Mutation Rates . . . . .	538

Checking for Contamination . . . . .	539
Radiation-Induced Mutants . . . . .	540
Gamma-ray Mutagenesis . . . . .	540
Gamma-ray Mutants . . . . .	541
Examples of Breeding Programmes Using Gamma-ray Mutagenesis . . . . .	542
Heavy-Ion Beam Mutagenesis . . . . .	543
Heavy-Ion Beam Mutants . . . . .	544
Chemical Mutagenesis . . . . .	545
Alkylating Agents . . . . .	545
Sodium Azide . . . . .	546
Other Types of Chemical Mutagen . . . . .	546
Identification of Mutants . . . . .	547
TILLING . . . . .	548
Screening for Mutants . . . . .	548
Results . . . . .	551
Causal Mutation for a Given Phenotype . . . . .	552
Insertional Mutagenesis . . . . .	553
Gene Silencing . . . . .	554
Site-Directed DNA Sequence Modification (Genome Engineering) . . . . .	554
Engineered Nucleases . . . . .	555
Mutagenesis . . . . .	555
Gene Replacement or Editing . . . . .	556
Gene Insertion (SDN-3 Technique) . . . . .	556
Site-Directed Structural Changes . . . . .	557
Rice Example . . . . .	557
CRISPR/Cas9 System for Plant Genome Editing . . . . .	557
Somaclonal Variation . . . . .	559
Mutation Breeding Versus Transgenesis . . . . .	559
<b>17 Genetically Modified Crops . . . . .</b>	<b>561</b>
Introduction . . . . .	561
Transformation Techniques . . . . .	562
Agrobacterium-Mediated Transformation . . . . .	563
Agrobacterium-Mediated Transformation of Potatoes . . . . .	564
Naturally Transformed Sweet Potato . . . . .	565
Biostatic-Mediated Transformation . . . . .	565
Biostatic-Mediated Transformation of Rice . . . . .	566
Clean-Gene Technology and Gene Stacking . . . . .	567
Regulations Governing GM Crops, Food and Feed . . . . .	568
EU Legislation . . . . .	568
EFSA Guidance . . . . .	569
USA Legislation . . . . .	571
Cisgenic, Intragenic and Transgenic Crops . . . . .	572

Gene Silencing . . . . .	573
Herbicide Tolerance . . . . .	574
Herbicide Tolerant Crops in the USA 1996–2012 . . . . .	575
Virus Resistance in Papaya . . . . .	576
Insect Resistance . . . . .	577
<i>Cry</i> Endotoxin Genes from <i>Bacillus thuringiensis</i> . . . . .	577
Insect Resistant Cotton in China . . . . .	578
Insect Resistant Cotton in India . . . . .	578
Drought Stress Response of Insect Resistant Maize in Europe . . . . .	579
Drought Tolerant Maize . . . . .	580
Transgenic Drought Resistance and Tolerance . . . . .	581
Reluctance to Accept Genetically Modified Potatoes . . . . .	582
Monsanto's Programme in the USA . . . . .	582
Amflora in Europe . . . . .	582
Late Blight Resistant Potatoes . . . . .	583
Simplot's Programme in the USA . . . . .	584
Cultivation of GM Potatoes in Centre of Origin and Diversity . . . . .	585
Low Lignin Alfalfa . . . . .	586
Golden Rice . . . . .	587
<b>18 Climate Change and Resistance to Pests and Diseases . . . . .</b>	<b>591</b>
Introduction . . . . .	591
Breeding for Resistance to Pests and Diseases . . . . .	592
Types of Resistance . . . . .	595
Qualitative Oligogenic Resistance: The Gene-for-Gene Concept . . . . .	596
Resistance to the Rust <i>Melampsora lini</i> in Flax: Flor's Gene-for-Gene Concept . . . . .	596
Examples from Other Crop-Parasite Interactions . . . . .	598
Physiological Races and Host Differentials . . . . .	600
<i>R</i> -Genes from <i>Solanum demissum</i> for Resistance to Late Blight of Potatoes . . . . .	600
Successes and Partial Successes . . . . .	602
Horizontal and Vertical Resistance . . . . .	603
Horizontal Resistance . . . . .	605
Screening for Resistance . . . . .	605
Late Blight in the Foliage of Potatoes . . . . .	606
Powdery Scab of Potatoes . . . . .	608
Gangrene of Potatoes . . . . .	608
Measurement of Resistance: Disease Progress Curves . . . . .	609
Theory . . . . .	609
Practice . . . . .	610
Disease Nurseries . . . . .	613
Examples of Relevance to Breeding for Durable Resistance . . . . .	615
A Further Look at Resistance to Late Blight of Potatoes . . . . .	616
Resistance of Stirling . . . . .	617

<b>Contents</b>		<b>xxiii</b>
Resistance of Sárpo Mira . . . . .	618	
Slow-Rusting Wheats . . . . .	619	
Breeding Slow-Rusting Wheats . . . . .	620	
Mlo Powdery Mildew Resistance in Barley . . . . .	621	
Broad Spectrum Bacterial Resistance . . . . .	622	
Durability of a Virus Resistance Gene . . . . .	623	
Durable Resistance . . . . .	623	
Farming and the Fate of Wild Nature . . . . .	624	
<b>Epilogue: Way Ahead . . . . .</b>	<b>627</b>	
<b>References . . . . .</b>	<b>635</b>	
<b>Index . . . . .</b>	<b>675</b>	



# Prologue

## Plants

### *Evolution of Plants*

Plant breeding is the genetic improvement of cultivated plants through human selection, a process which began when plants were first brought into cultivation for human use, as early as 13,000 years ago (Balter, 2007). The plant species available for cultivation were the products of millions of years of evolution by natural selection. Our understanding of the history of plant life on Earth comes from an examination of the fossil record of plant remains and biomolecules, complemented with analysis and classification of the patterns of morphological and genetic variation among present-day plants (phylogenetics). The timescale is estimated from geological records and from the rate of production of genetic differences seen today (the molecular clock of phylogenetics). Detailed accounts of plant evolution, together with theories of the mechanisms driving the process, can be found in books such as *The Evolutionary Biology of Plants* by Niklas (1997), the beautifully illustrated *Green Universe* by Blackmore (2012) and *The Evolution of Plants, Second Edition*, by Willis and McElwain (2014). Here it is sufficient to list and briefly comment on some key events dealt with by these authors:

1. The evolution of the first forms of life (single prokaryotic cells from protocells with membranes that enclose the chemical reactions that form the basis of life), possibly in tidal pools, by 3.5 billion years ago. At that time, carbon dioxide concentrations were between 30 and 100 times higher than present levels of 360 ppm, global temperatures could have been between 30 and 50 °C (high enough to sustain liquid water on Earth), and oxygen accounted for <0.01 % of total atmospheric concentration.
2. The origin and development of photosynthetic cells; in particular, cyanobacteria (blue-green algae) capable of using light energy from the sun, water and carbon dioxide to synthesise complex sugars (with the release of oxygen) some

- 2.7 billion years ago. About 2.5 billion years ago, oxygen levels rose significantly to around 2.1 % and remained at that level for the next 2.0 billion years.
3. The formation by endosymbiosis of eukaryotic cells with a nucleus and organelles (mitochondria resembling purple non-sulphur bacteria for energy production and chloroplasts resembling cyanobacteria for photosynthesis); in particular, unicellular green algae approximately 1.87 billion years ago.
  4. The acquisition of a multicellular construction; for example, green algae living in an aquatic medium some 800 million years ago (mya), that is, 800 megaannums (Ma).
  5. The colonisation of land (and air) by plants from 470 Ma. The initial colonisers were relatively small (<1 m in height) plants; restricted to damp, moist regions; and related to our living bryophytes (mosses, liverworts and hornworts). The atmosphere was very rich in carbon dioxide, between 8 and 15 times higher than present, but still low in oxygen at 4 %. The global average surface temperature was thought to be approximately 21 °C.
  6. The evolution of (seedless) vascular plants with conducting tissues for the transport of water (lignified) and food from 430 Ma. The result was large plants filling every ecological niche available, including spore-producing trees from 390 Ma (giant horsetail tree, giant Lepidodendron tree, tree-ferns and the Archaeopteris tree with its particularly advanced vascular system, true roots and megaphyllous leaves). By 360 Ma forests had widespread global distribution. Between 360 and 299 Ma atmospheric carbon dioxide levels rapidly declined from 3600 to 300 ppm, probably as a result of the global expansion of vascular plants. In contrast, there was a remarkable increase in oxygen to above 21 % (current level), with a peak of between 28 and 30 % around 280 Ma.
  7. The evolution of naked (fruitless) seed plants (gymnosperms), initially trees, from 350 Ma, with a major radiation of new groups of plants from 290 to 250 Ma (including cycad and ginkgo trees whose pre-pollen microspores produce motile male gametes), and of conifers (cone-bearing and airborne pollen producing) from 250 to 200 Ma. From 299 to 250 Ma, carbon dioxide levels increased from 300 to 1500 ppm and then fluctuated until 140 Ma. Today the tallest known trees at 115 m are coast redwoods (*Sequoia sempervirens*) from California, and the largest due to their girth are giant redwoods (*Sequoiadendron giganteum*) at 84 m and 1486 m<sup>3</sup> of timber.
  8. The evolution of the flowering plants (angiosperms) from 140 Ma, the first of which were probably herbaceous, weedy, small shrubs in low latitudes, but could have been aquatic plants. Angiosperms have seeds which are fully enclosed within a new and specialised plant organ, the carpel. Although the most rapid interval of angiosperm radiation was around 100 Ma, the grasses (Poaceae) did not emerge until 70 Ma. Initially they occupied an understory habitat within closed forested vegetation, but by 20 Ma, probably as a result of increasing global aridity, they had become the dominant group of plants in truly open habitat grasslands in many areas, ecosystems that we now describe as tropical and subtropical savannah, temperate grasslands and steppes. Today grasses provide over 50 % of the carbohydrates in our human diet.

Our present-day, angiosperm-dominated, flora is the result of major changes in the overall composition and distribution of global vegetation over the past 66 million years, associated with major extrinsic environmental change. During this period tectonic processes occurred in the Earth's crust, as a result of which the continental plates moved into their present positions, today's prominent mountain ranges formed and new ocean current systems were established. An initially warm ( $>30^{\circ}\text{C}$ ) global climate became increasingly cool (towards  $15^{\circ}\text{C}$  global mean) and arid, culminating in a build-up of ice at the poles, greater temperature gradients from equator to poles and the formation of deserts. Carbon dioxide concentrations fell from about 1200 ppm to around 300 ppm. From about 2.5 Ma recurrent glacial-interglacial cycles have occurred, closely linked to variations in the incoming solar radiation. These result from changes in the Earth's orbit around the sun and the tilt and wobble of the Earth on its axis. Concentrations of atmospheric carbon dioxide fell to averages of around 180 ppm during the glacial periods compared with 300 ppm during interglacial periods. During glacial periods temperate vegetation became isolated in regions where microenvironmentally favourable conditions existed and then re-expanded from these regions during interglacial periods.

Biogeographical maps of the Earth's flora can be found in the book by Willis and McElwain (2014), but we do not need to consider them further. Before starting the history of plant breeding, however, it is worth pausing for a moment to reflect on the extent to which we need plants for our well-being, as this provides the context for their improvement through breeding.

## ***Use of Plants***

People use plants for many purposes as seen in Economic Botany Collections such as the ones at the Royal Botanic Gardens, Kew (1998), which now consist of over 76,000 specimens. A few examples will suffice. Plants provide us with basic food in the form of carbohydrate and protein, as well as essential fatty acids, vitamins and minerals. Carbohydrates come from cereals (wheat, rice and maize), tubers (potatoes) and roots (cassava) and proteins from legumes and their pulses (beans, peas and lentils). Vegetable oils for both food and industrial use come from the crushed or pressed seeds of temperate crops such as oilseed rape and linseed, Mediterranean ones like sunflowers and olives and tropical ones like peanuts and palm oil. Plants also provide fodder for our livestock, including grass for ruminants as they can digest the cellulose in plant cell walls. We flavour our food with herbs and spices and sweeten it with sugar, both cane and beet. We make wine from grapes and beer from barley, while tea, coffee and chocolate are popular beverages. We clothe ourselves with plant fibres such as cotton, coloured with dyes like indigo in blue jeans, and use plants for hats, shoes and adornments such as seed necklaces. We are clean and sweet smelling thanks to scented soap made, for example, from palm oil and lavender. We still rely on plants for medicines. The latex of poppy capsules provides morphine and codeine for pain relief, and for many years, quinine from

cinchona bark was the best treatment for malaria. Indeed, it can be argued that the settlement of white Europeans in Africa, with its political, economic and social consequences, was only possible through quinine (Hobhouse, 1992). Currently one of the most effective treatments for malaria is artemisinin (qinghaosu in Chinese), a drug extracted from the herb *Artemisia annua* (sweet wormwood), which had been used in traditional Chinese medicine for many centuries before being rediscovered by the Western world in the 1970s (Benson, 2012).

Plants provide us with fuel, traditionally wood to burn for heating and cooking, but increasingly as a fuel for electricity generation together with grasses such as *Miscanthus × gigantus*. Today wood harvesting systems include short rotation coppice, for example, willow and poplar; and since the mid-1800s paper has been made from wood pulp, having first been made by the Chinese around 100 BCE (before the Common Era) from plant fibres such as hemp. Whereas wood for fuel makes use of the current growth of trees, fossil fuels (coal, oil and natural gas) are a finite resource from past trees and vegetation. Biodiesel from oilseed crops and bioethanol produced by fermenting plant sugars are modern alternative fuels to diesel and petrol, respectively. Second-generation biofuels are anticipated in which cellulose cell walls are also converted to fuel. Plants also provided the materials for water transport and now rubber for the tyres of modern vehicles. Musical instruments, sports and games equipment and storage containers (baskets and boxes) have all been made from plants, as have traps and nets for hunting and fishing. Finally, we find plants beautiful and use them for decorative purposes.

More details on our uses of plants can be found in *Why People Need Plants* edited by Wood and Habgood (2010). We should, however, never forget that we rely on plants as a whole to capture the sun's energy to make the food we eat and to produce the oxygen we breathe and, perhaps most importantly of all, that photosynthesising organisms have shaped and still shape the Earth's climate.

# **Part I**

## **Historical Introduction**

# **Chapter 1**

## **Domestication, Dispersion, Selection and Hybridization of Cultivated Plants**

### **Introduction**

Plant breeding began when plants were first brought into cultivation for human use, as early as 13,000 years ago in the Near East, and subsequently and independently in many different parts of the world (Balter 2007). It is estimated that some 2500 plant species have been domesticated by humans worldwide, with over 160 plant families contributing one or more crop species (Meyer et al. 2012). In contrast, the dominant plants in the world today, the angiosperms, comprise 457 families and between 270,000 and 400,000 species (Willis and McElwain 2014).

### ***Involvement of Humans***

Between 75,000 and 13,000 years ago human (*Homo sapiens*) migrants from Africa had colonized much of the world, during a period of particularly unstable climatic conditions (Murphy 2007). Furthermore, they had developed the technical (use of hands) and intellectual (use of brain) abilities required for a completely new way of life involving the growing of crops and animal husbandry, the so-called Neolithic Revolution (Hahlbrock 2009). But they could easily satisfy their dietary requirements through the hunting of animals and the gathering of food from wild plants. So why did these hunter-gatherers become farmers? There probably isn't one simple answer that applies to all of the different situations where plants were brought into cultivation, and the transitions probably occurred gradually at different rates in different places. Nevertheless, recent advances of knowledge in a number of disciplines allow convincing scenarios for some areas including the Near East (part of southwest Asia), particularly the coastal regions at the eastern end of the Mediterranean Sea and the area around the Tigris and Euphrates rivers (Murphy 2007).

## ***Transition to Farming in the Near East (Southwest Asia)***

The earth's last glacial period occurred from 110,000 to 11,500 years ago and was at its maximum from 25,000 to 15,000 years ago (Murphy 2007). Overall the period was appreciably cooler and drier than our climate today, and saw thick ice sheets develop across most of northern Europe and Canada, while further south prairie-like grassland replaced lush forests. However, the period was also characterized by great variations in the global climate with sudden cooling and warming periods, such as the ones that ended the glacial period. There was a sudden rise in temperature between 14,680 and 14,600 years ago, followed by a slow fluctuating fall between 14,600 and 12,800 years ago which included more rapid cooling for the last 100 years, and a final sudden rise between 11,640 and 11,620 years ago. This was the beginning of the relatively congenial interglacial period known as the Holocene Era which has continued to the present day. Around 14,000 years ago the climate in most parts of the world was at least as warm and moist as today. The cold, dry period between 12,800 and 11,640 is known as the Younger Dryas Interval, or the mini Ice Age. The interested reader can get estimates of actual temperature changes from the article by Peteet (2000). All of the changes mentioned had dramatic consequences for animal and plant life, and hence the types of food available for hunter-gatherers. In the Mediterranean Basin and Near East, from about 50,000 years ago, changes in the eating habits of human populations occurred, with a shift in hunting from large to small animals and an increasing use of plants. Initially all kinds of plants were used but the period from 25,000 years ago saw a more extensive use of wild grasses, including members of the cereal family with their larger, starchy seeds.

A rare insight into the plant foods being gathered 23,000 years ago comes from the Stone Age site Ohalo II, on the southwestern shore of the Sea of Galilee in Israel (Piperno et al. 2004). Excavations at the site revealed the remains of a camp which included six brush huts built from branches and leaves, hearths and a human grave; thus confirming the semi-sedentary nature of the community of hunter-gatherers. A collection of more than 90,000 plant remains was recovered from the site (Weiss et al. 2004). Besides cereals (wild emmer wheat and barley), small-grained grasses made up a large component of the remains. Sites from later periods revealed that over the next 15,000 years small-grained grasses were gradually replaced by the cereals and ultimately disappeared from the diet. Analysis by Piperno et al. (2004) of starch grains recovered from a ground stone artefact at Ohalo II, provided the earliest direct evidence for human processing of grass seeds, including barley and possibly wheat. Associated evidence for an oven-like hearth was also found at the site, suggesting that dough made from grain flour was baked. In other words, foragers living in simple huts at Ohalo II were able to pound and grind hard, fibrous seeds into easily digestible foodstuffs, which they could also bake.

During the warm interlude from 15,500 years ago, forests and woodlands became re-established in the Near East and lessened the reliance of the inhabitants of the area on cereals, although dense stands of large-grained wild grasses would

have been available at certain times of the year. An interesting account can be found in the review by Bar-Yosef (1998) of the emergence and nature of a cultural group called the Natufians in the Mediterranean Levantine area of the Near East, with its hunter-gathering but sedentary lifestyle based in semipermanent villages. Bar-Yosef (1998) provides convincing evidence that the Natufians used sickle blades (glossy pieces) for harvesting cereals, with the advantage over beaters and baskets of maximizing the yield from a limited area. He sees the emergence of farming communities as a response to the effects of the Younger Dryas on the Late Natufian culture in the Levantine Corridor. However, to date, the earliest example of the deliberate cultivation of cereals is 13,000 years ago at the prehistoric settlement of Abu Hureyra (35.529°N, 38.249°E) on the south side of the Euphrates Valley, 130 km east of the present-day city of Aleppo (Hillman et al. 2001).

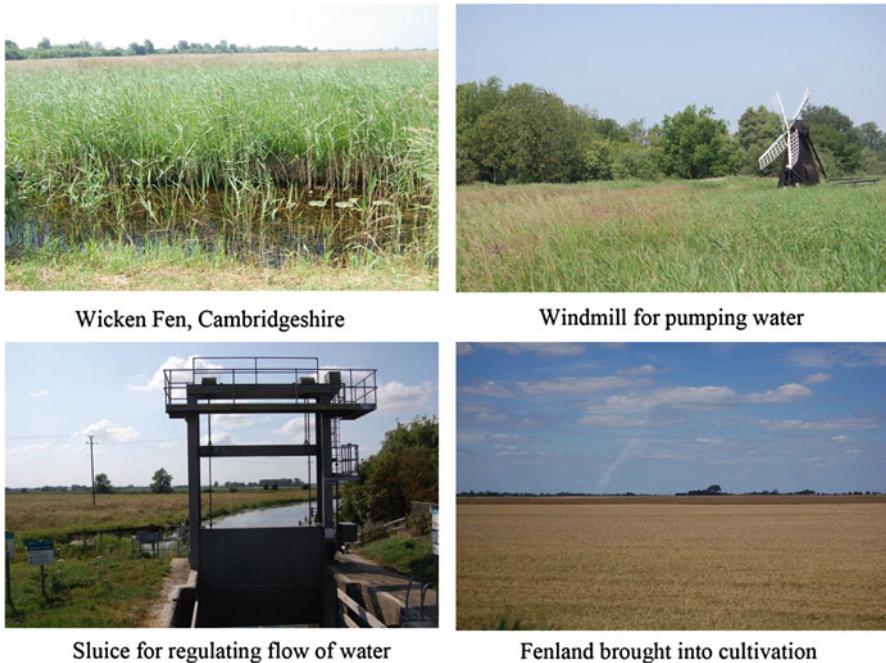
So far, Abu Hureyra provides the clearest, continuous record anywhere in the world of an *in-situ* transition from foraging to farming, and it is well worth reading in full the paper by Hillman et al. (2001). Briefly, their well argued conclusions, from the vast numbers of plant remains examined from different occupation times, are as follows. When hunter-gatherers first settled at Abu Hureyra around 13,200 years ago, they were easily able to forage more than 100 species of edible seeds and fruits, and they occupied the site year round. Then after 300–400 years, a rapid slide into aridity occurred over the whole area, coinciding with the Younger Dryas climatic episode. The result was a dramatic succession of losses of staple wild foods. The solution adopted by the settlers to this problem was to cultivate certain of their wild cereals in areas with enhanced soil moisture, which they had cleared of the natural dense scrub. An annual wild rye was definitely cultivated, either *Secale cereale* L. subsp. *vavilovii* (Grossh.) Zhuk. or *S. iranicum* Kobyl., and probably also a wild wheat. Later lentils and other large-seeded legumes were cultivated. It was these beginnings of cultivation some 13,000 years ago that gave rise to the integrated grain-livestock Neolithic farming systems of the early Holocene. The transition from exclusive use of wild foods to dependence on cultigens took about 2500 years. Once established, farming sustained much greater populations, not all of whom needed to be involved in food production, because farmers obtain far greater productivity than hunter-gatherers in terms of food calories per unit area of land. Accounts of other areas and situations where plants were brought into cultivation can be found in the book *People, Plants & Genes* by Murphy (2007).

## ***Bringing Land into Cultivation***

The history of plant breeding is intimately associated with bringing land into cultivation to increase food production, a process that has continued to the present day. In contrast, it is only in the last 100 years that there has been a dramatic increase in yields per unit of land, and an accompanying rapid growth in human population size. The total area of land for agricultural production in 2011 was

thought to be between 1.22 and 1.71 billion hectares (International Institute for Applied Systems Analysis-IIASA). McMahon (2013) summarized the work done by IIASA on current and future land use as follows. The IIASA model shows that about a quarter of the world's land mass (excluding Antarctica) is not productive, with 22 % occupied by desert, mountain, inland lakes or rivers and 3 % used for human settlement and infrastructure. Another 28 % of the land is under closed forest. About 35 % is covered by grassland or open woodland, much of which is used by pastoralists for grazing livestock. This means that about 12 % of the planet's land mass, or 1.6 billion hectares, is cultivated for agriculture. IIASA estimated that 3.1 billion hectares of additional, uncultivated land is agro-ecologically suitable for rain-fed crop production. However, if protected areas such as national parks and also forests are excluded, we have 1.3 billion hectares of grassland and open woodland suitable for agricultural expansion, of which 60 % is in South America and Africa. Although in a sense reassuring for future needs, McMahon (2013) argues that bringing this land into production would pose an enormous social, economic and environmental challenge. Furthermore, there are concerns about degradation of current agricultural land, as we shall see in the next chapter. If possible, it would be better to get more food out of existing farmland; a topic we also need to examine in the next chapter. IIASA and the United Nations' FAO expect this to happen and therefore predict that the amount of cultivated land in the world will only grow by 10 % between now and 2050.

Bringing land into cultivation has been contentious when it involved disputes over ownership of land, and it has certainly not always been peaceful, as I know from my local history. The period from 1600 to 1900 saw 7000 km<sup>2</sup> of wet fens in Eastern England brought into cultivation in the form of privately owned and managed fields, by drainage, clearance and enclosure (Rotherham 2013). This was followed from the 1950s by the development of intensive, petrochemically driven arable farming; an aspect of modern industrial farming (Fig. 1.1). Earlier attempts at reclamation by the Romans, and then by monastic communities, had only limited success because it was not sufficient to cut peat and to build dykes. Sluices were needed to control the flow of water which was pumped from low levels into higher drainage channels that flowed to the sea; and also to prevent inundations from the sea. Initially the power came from windmills; later from steam, oil and electricity. The 'Great Drainage' was started in 1626 by King Charles I who employed the skills of Dutch engineers and settlers. It was deeply resented by the local people who supported Oliver Cromwell and his Parliamentarians in the English Civil Wars (1642–1651) that resulted in the King being tried and executed in 1649, and a period of 11 years without a monarch. When the monarchy was restored, it was by 'Consent of Parliament'. A consequence of the 'Agricultural Improvement' was the loss of fens rich in fish, fowl and deer; the loss of peat for fuel and sedge for thatching and kindling; and the loss of the livelihoods of local people who were driven into cities. Nature reserves like Wicken Fen provide a glimpse of what has been lost. Ironically in Eastern England today new wetlands are required to store water for summer irrigation and to avoid the worst of problems from winter and summer floods. In other words, there is a need to align modern



**Fig. 1.1** Bringing wet fens in Eastern England into cultivation

intensive farming with more sustainable flood defences, water supply, nature conservation and a vibrant local economy. Today it is recognized that globally, wetlands (6 % of Earth's surface) play an important role in maintaining the equilibrium of the water cycle and preventing flooding, as well as having an economic importance for the populations of people living in or around them (Benson 2012). Likewise, globally it is recognized that other kinds of habitats, such as rain forests, are equally important in maintaining the world's major ecosystems.

The Great Plains of North America provides an example of the rapidity with which land can be brought into cultivation (Wikipedia). In 1800 they were dry prairies of deep-rooted grasses such as switchgrass (*Panicum virgatum*). These were grazed by buffalo (bison) which were hunted by the Native Indian population. One hundred years later, a railroad crossed the Great Plains and the buffalo had been hunted to extinction for their hides by white settlers from the east coast. The land was now farmed by settlers encouraged to do so by the US Homestead Act, signed by President Abraham Lincoln in 1862. They prepared the land with steel ploughs, grew cereal crops of wheat and maize, and irrigated them with water raised by wind pumps from a vast natural underground aquifer (The Ogallala). They also reared cattle on watered grassland enclosed by barbed wire fences. The result was a tremendous technological achievement, but destroyed the way of life of the local indigenous people who became confined to reservations.

Local people have no doubt resented the land deals that have occurred in developing countries following the 2008 food crisis and concerns about food security. McMahon (2013) explains that these deals target poor countries where land is cheap or can be obtained for free. So far 358 such deals have been identified in Africa covering 15 million hectares. They usually involve a transfer from government or local communities to a foreign company in the form of a long-term lease. The goal is to effect a transformation in how the land is used; for example, by ploughing up grassland or clearing forest, and planting crops. Often, the ultimate goal is to integrate land that has been used for local subsistence into global commodity supply chains. The term ‘land grab’ is used because of poor recognition of the land rights’ of local indigenous people. The investors range from governments of food importing countries (for example, oil-rich Arab states) worried about security of supply, to financial investors (pension funds, insurance companies and wealthy families) excited by the profits to be made from high food prices. In between there are global corporations looking for profit as well as security of supply, for example, in tropical plantation crops. Today, as always, there is a political, economic, social and environmental context to plant breeding.

## Domestication

### *Centres of Diversity and Origin*

Our modern view of the origin, geography and evolution of cultivated plants dates to the work done from 1920 to 1940 by Nickolay Ivanovich Vavilov (1940), who did so much to promote the value of scientific plant breeding in feeding the world’s growing population. It was therefore both ironic and tragic that he died of starvation in prison in Stalin’s Soviet Union in 1943 (Love 1992).

Vavilov considered knowledge of the origin of a cultivated species and the route of its geographical dispersal to be of key importance for understanding the course of its evolution. His extensive investigations were based on wide-ranging expeditions by Soviet botanists and agronomists, which covered the major part of the world’s agricultural areas, and which resulted in the collection of more than 200,000 samples. These were used for a multifarious study of the living material which was sown or planted under various conditions. Vavilov considered 1000 of the most important cultivated plant species which then occupied over 99 % of the world’s cultivated land. His conclusions were based on the centre(s) of diversity of the cultivated species, data and inferences on the occurrence of ancestral species and wild forms (types), as well as evidence from archaeology, history and linguistics (names of a cultivated plant in different languages).

He considered eight, but recognized seven main geographical centres of origin of cultivated plants that were associated with fertile river valleys and plains, and also with ancient agricultural civilizations:

1. South-Asiatic tropical centre associated with high-level Indian and Indo-Chinese civilizations.
2. East-Asiatic centre associated with ancient Chinese civilizations.
3. Southwestern-Asiatic centre associated with ancient civilizations of Iran, Asia Minor, Syria and Palestine
4. Mediterranean centre associated with Etruscan, Hellenic and Egyptian civilizations.
5. Abyssinian centre associated with Abyssinian civilization.
6. Central American geographical centre associated with Mayan civilization.
7. Andean centre associated with pre-Incan and Incan civilizations.

Meyer et al. (2012) have also made the point that patterns of domestication activity in the different centres are extremely variable over time, but that peaks coincide with the rise of major civilizations. The civilization that lasted the longest (3300–330 BCE) was the Egyptian one based on the natural flooding of the Nile, which provided both water and rich silt each July to September from its source in the Ethiopian highlands (Bell 2009). A large Egyptian workforce extended the natural delta with a sophisticated irrigation system that provided enough cultivated land to feed the Roman Empire with wheat. Eventually the water supply was insufficient, and this should be a warning to us all today. Interestingly, Rome used aqueducts to bring water from mountain sources for drinking and washing, but not for agriculture. The link between the close proximity of food production and the city had been broken, a process which had been started by the Greeks in Athens. Our large modern cities depend on piping in clean water, removing human excrement through sewers, and providing food from a global marketplace.

Today as many as 11 or 12 centres of origin are recognized (Balter 2007; Hahlbrock 2009), with various evidence for 24–28 separate regions where crop domestication occurred independently (Purugganan and Fuller 2009; Meyer et al. 2012). Domestication occurred in these centres from about 13,000–10,000 years ago to as recently as 4000–3000 years ago. Our most important food crops come from different centres, a reflection of the wild species that were present in those areas: wheat and barley from Vavilov's Southwestern-Asiatic centre, rice from his South-Asiatic tropical centre, maize from his Central American centre, sorghum from his Abyssinian centre and potatoes from his Andean centre (and also probably cassava from South America). Hancock (2012) points out that these starchy staples were complemented with domestication of a high protein vegetable crop (legume) and a fiber crop.

Vavilov's views on the direct relationship between centres of diversity and centres of origin are now considered too simplistic. But the collection and conservation of germplasm of cultivated species and their wild relatives in genebanks was a direct outcome of his work. A recent example was the opening on 26 February 2008 of the Svalbard Global Seed Vault, which is located for security deep in a remote mountainside close to the North Pole.

## ***Features of Domestication***

The archaeological evidence now available on the transition from hunter-gatherers to agriculturalists favours a protracted transition rather than a rapid one; as summarized by Allaby (2010) in his review of domestication. He points out that under the protracted transition paradigm, numerous interacting factors need to be taken into account in order to understand the evolution of the process and the resulting patterns of genetic diversity. Hence we face a complex problem and should not be surprised when we find more than one interpretation in the literature. Meyer et al. (2012) explored patterns and processes in crop domestication through an historical review and quantitative analysis of 203 global food crops. Whilst it is not possible to present all of their results ([www.cropdomestication.com](http://www.cropdomestication.com)), a number of key points are as follows.

1. There was good evidence of multiple origins for only 38 (19 %) of the crops analyzed, that is where domestication occurred independently from the same ancestor(s) in different locations or times. The majority of crops appear to have been domesticated once from their wild ancestor(s), followed by dispersal.
2. Domesticated annuals are most prevalent in regions of arid climate, biennials in semiarid climates and trees in humid climates, whereas non tree perennials are relatively evenly distributed. These results are consistent with the expected proportions of such life cycles in the regional flora.
3. The suite of traits that marks a crop's divergence from its wild ancestor(s) is defined as the 'domestication syndrome'. It may include combinations of several different traits such as seed retention (non shattering), increased fruit or seed size, changes in branching and stature, changes in subterranean organs, changes in life cycle (annual, biennial or perennial) and reproductive strategy (self- or cross-pollination or vegetative), changes in ploidy and changes in secondary metabolites. Eighty-four per cent of the crops had between two and five domestication syndrome traits, while some had as many as seven; the average was 2.8. The most common change has been in secondary metabolites affecting, for example, toxicity, colour and flavour.
4. The average time span over which domestication occurred (i.e. for the domestication syndrome to be fixed in the crop) was 3767 years for trees and 2638 years for annuals; but there was a clear, progressive decrease in mean time to completion of domestication, in particular between crops domesticated over 8000 years ago and more recent crops, with kiwi, cranberry and grapefruit providing examples of crops domesticated in the last 200 years.
5. The appearance of domesticated annuals increased from 9000 to 4000 years ago and then started to slow. The two periods of sharp increases in the cultivation and domestication of trees and non tree perennials, 6000 and 3000 years ago, coincided with innovations in vegetative propagation through cuttings and scion grafting, respectively. The first domesticated biennials appeared only in the last 6000 to 5000 years, and their rise between 3000 and 1000 years ago corresponded with the peak of trade and activity in the Roman Empire throughout the Mediterranean, where many biennials were domesticated.

## ***Continuing Debates About Domestication***

Meyer et al. (2012) found major gaps in the literature regarding the place of origin, time of domestication and identity of wild ancestor of crop species, and concluded that just 30 % of species examined had high confidence scores for availability and quality of data. Ten per cent of the crops studied were under dispute with regard to their origins or wild ancestor. Not surprisingly then, new discoveries in archaeology, botany (e.g. analyzing microscopic plant remains such as starch grains) and molecular biology are still adding to our knowledge of the domestication and evolution of cultivated plants. Six examples will briefly be considered.

**Wheat:** A lively and inconclusive exchange of letters took place in *Science* in July 2006 (Kavanagh 2006) on whether or not it took over a millennium to domesticate wild wheat (*Triticum–Aegilops* group); that is, how quickly and effectively early farmers selected the initially rare variant that produced indehiscent ears with non-shattering spikelets (the ‘domestic spikelet’). It centred on an assessment and interpretation of the proportion of domestic spikelets in charred samples from four archaeological sites, located in northern Syria and southeastern Turkey, and dating from between 10,200 and 6500 years ago.

**Rice:** The history of rice (*Oryza sativa*) domestication has also been the subject of much debate. According to the summary in Virmani and Ilyas-Ahmed (2007), the genus *Oryza* probably originated 130 MYA (million years ago) and spread as a wild grass in Gondwanaland, the supercontinent that broke up into our present day ones. The two cultivated species *O. sativa* (Asian rice) and *O. glaberrima* (African rice) originated from a common ancestor, a wild perennial grass, and underwent parallel evolution. The wild progenitor of *O. sativa* is the common Asian wild rice *O. rufipogon* which varies from perennial to annual types. The latter, called *O. nivara*, was domesticated and resulted in *O. sativa*. Some authors favour the hypothesis that the two main subspecies (or variety groups) of Asian rice (*indica* and *japonica*) arose from genetically distinct gene pools within the common wild ancestor, *O. rufipogon*, followed by limited transfer (introgression) of key variants (alleles) of domestication genes between the divergent gene pools (Kovach et al. 2007). This explains why some major domestication alleles are common to all cultivars of *O. sativa*. Other authors consider that for both Asian and African rice, domestication was a single event, the selection of a non-shattering gene, followed by evolution that led to their genetic complexity (Vaughan et al. 2008). Support for a single origin of Asian rice has come from demographic modelling by Molina et al. (2011) based on sequencing 630 gene fragments in a diverse set of wild and domesticated rice accessions. These authors also used their molecular data to date the origin of domestication at 8200–13,500 years ago. This is consistent with known archaeological data that suggests rice was first cultivated at around this time in the Yangtze Valley of eastern China. Further support for a single domestication comes from the work of Huang et al. (2012a). They surveyed sequence variation at 42 genome-wide sequence tagged sites (STS) in 108 *O. rufipogon* accessions from

throughout the native range of the species. They concluded that there are two genetically distinct *O. rufipogon* groups, Ruf-I and Ruf-II. The former was found mainly in China and the Indochinese Peninsula, and showed genetic similarity with cultivated *indica* rice. In contrast, Ruf-II, mainly from South Asia and the Indochinese Peninsula, was not closely related to cultivated rice. Furthermore, *japonica* rice was not similar to either *O. rufipogon* groups, although a *japonica*-like component was found in a few Ruf-I individuals, mainly from China. A domestication scenario consistent with all these results is that a primitive cultivar was originally domesticated from the Ruf-I group in China, and both *indica* and *japonica* rice were its descendants. In contrast, Huang et al. (2012b) compared genome sequences from 446 geographically diverse accessions of *O. rufipogon* with 1083 cultivated *indica* and *japonica* cultivars. They concluded from the patterns of variation at eight million SNP (single nucleotide polymorphism) sites that *japonica* rice was first domesticated from a specific population of *O. rufipogon* around the middle area of the Pearl River in southern China (Guangxi province). The major groups, aromatic, temperate *japonica* and tropical *japonica* were derived from this domesticated *japonica* (see next paragraph). *Indica* rice (major groups, *indica* and *aus*) was subsequently developed from crosses between *japonica* rice and local wild rice as the initial cultivars spread into South East and South Asia. Hence rice may have arisen in one geographical region of Asia from where it spread with farming to become the major food crop for much of the world's population. Huang et al. (2012b) were also able to discover and analyze 55 domestication-related genetic loci.

Recently Wang et al. (2014a) have demonstrated that *O. sativa* germplasm can be further classified into five previously recognized major groups (*indica*, *aus*, aromatic, temperate *japonica* and tropical *japonica*) and a newly recognized one (*rayada*). They used 84 nuclear microsatellite markers to genotype a panel of 153 global rice cultivars and 826 cultivars (747 landraces and 79 modern cultivars) representing the diversity of Chinese rice germplasm. Furthermore, they identified three major groups (*indica*, temperate *japonica* and tropical *japonica*) in the Chinese germplasm and showed that Chinese temperate *japonica* contained higher diversity than that of the global samples. All seasonal, drought-tolerant and endosperm types of rice occurred within each of three major groups of Chinese cultivars. The other cultivated species of rice, *O. glaberrima*, comes from tropical West Africa and has the weedy annual *O. barthii* as progenitor (Guimarães 2012). Cultivation of this African species is confined to its region of origin. Hybridization between Asian and African rice is made difficult by post zygotic barriers that result in the hybrid being pollen sterile and only partially female fertile (Garavito et al. 2010).

*Maize:* The recovery of 7100-year-old maize (*Zea mays*) pollen from San Andrés in Mexico provided evidence for earlier domestication than once thought (Piperno 2001). Subsequently Piperno et al. (2009) reported starch grain and phytolith (siliceous plant remains) data from a site in the Central Balsas Valley of Mexico that indicate that maize was present 8700 years ago, and was domesticated in the seasonal tropical forest of southwestern Mexico rather than in the semiarid

highlands. Molecular marker data confirmed that maize originated from annual teosinte (*Zea mays* subspecies *parviglumis*) around 9000 years ago in the mid to lowland regions of southwest Mexico where subspecies *parviglumis* grows endemically (van Heerwaarden et al. 2011). The similarity of the cultivated maize of the highlands to subspecies *parviglumis* probably reflected admixture with another annual teosinte, *Zea mays* subspecies *mexicana*.

**Sunflower:** Intriguingly, domesticated sunflower (*Helianthus annuus*) remains which date to 4700 years ago were also found at San Andrés, suggesting a Mexican origin of sunflower rather than an eastern North American one (Piperno 2001). However, further research by Harter et al. (2004) supported the hypothesis that extant domesticated sunflowers arose from wild populations in central USA via strong genetic drift, as would be expected from a selective bottleneck during domestication in eastern North America. They used model-based methods and data from 18 microsatellite loci distributed across the sunflower genome. They evaluated genetic relationships and reconstructed the pattern of genetic drift among 21 populations of wild *Helianthus annuus* and eight Native American landraces from the USA and Mexico, as well as two modern cultivars (USDA and Mammoth). Nevertheless, an independent domestication in Mexico cannot be ruled out.

**Potato:** The cultivated potato (*Solanum tuberosum*) probably has more wild species relatives than any other crop, with around 200 tuber-bearing *Solanum* species distributed from the south-western United States (38°N) to central Argentina and adjacent Chile (41°S). Not surprisingly, the progenitors of cultivated potatoes have been the subject of much discussion. Spooner et al. (2005) have provided molecular taxonomic evidence for a single domestication in the highlands of southern Peru from the northern group of members of the *Solanum brevicaule* complex of (diploid) species. This group contains species such as *S. canasense*, *S. multidissectum* and *S. bukasovii*, some of which are not always clearly resolved and perhaps could be better reduced to a single species, *S. bukasovii*. The result of domestication was a diploid cultigen *S. tuberosum* Group Stenotomum (Dodds 1962) from which all of the other cultivated potatoes was derived. Group Stenotomum potatoes remained confined to the central Andes of Peru and Bolivia. In contrast, their tetraploid (four instead of two sets of chromosomes) descendants became grown throughout the upland Andes of South America (Group Andigena), and a secondary derivative of these became adapted to the long days of coastal Chile (Group Tuberorum).

**Common bean:** The common bean (*Phaseolus vulgaris*) provides an example of a debate about the origins of the ancestor of a domesticated species that appears to have been resolved by Bitocchi et al. (2012). The current distribution of the wild common bean encompasses a large geographical area from northern Mexico to northwestern Argentina. One hypothesis is that a population from the western slopes of the Andes in northern Peru and Ecuador was the ancestral population from which *P. vulgaris* originated, and then spread both northwards (Colombia, Central America and Mexico) and southwards (southern Peru, Bolivia and Argentina) to form the

Mesoamerican and Andean gene pools, respectively. The Mesoamerican origin of the common bean is an alternative and older hypothesis. Bitocchi et al. (2012) investigated the nucleotide diversity at five gene loci of a large collection of wild common bean accessions from the Mesoamerican (49 accessions) and Andean (47) gene pools and genotypes from northern Peru–Ecuador (6) that were characterized by the ancestral type I phaseolin. Their data clearly indicated a Mesoamerican origin of the common bean and subsequent dispersion in South America. The data strongly supported the occurrence of a bottleneck during the formation of the Andean gene pool. Furthermore, the results implied that separate gene pools in northern Peru–Ecuador and the rest of the Andes originated through different migration events from the Mesoamerican populations that were characteristic of central Mexico. Overall, Mesoamerica and the Andes are still recognized as the two major ecogeographical gene pools from which two independent domestications occurred, with the two wild and the two domesticated gene pools characterized by partial reproductive isolation. Mamidi et al. (2013) estimated that the two current wild gene pools diverged from a common ancestor about 111,000 years ago; based on sequence diversity at 13 loci for 24 wild genotypes from throughout the geographic range of the species. Subsequently, each gene pool underwent a bottleneck which resulted in population sizes in Mesoamerica and the Andes that were 46 % and 26 %, respectively, of the ancestral population sizes. The data of Bitocchi et al. (2012) also revealed the very complex geographical structure of the genetic diversity in Mesoamerica, with central Mexico and the Transverse Volcanic Axis, which originated about 5 MYA in the Late Miocene, as the cradle of diversity of *P. vulgaris*. Mesoamerican germplasm would appear to be a rich source of diversity for use in breeding programmes. Bitocchi et al. (2013) went on to compare 112 domesticated accessions (64 Mesoamerican and 48 Andean) with their 102 wild ones using data from the same five gene loci. They found a reduction in genetic diversity in both gene pools which was threefold greater in Mesoamerica compared with the Andes. However, as already mentioned, a major reduction in diversity had already occurred in the Andes. Their results highlighted a single domestication event within each gene pool, with the Oaxaca valley in Mesoamerica, and southern Bolivia and northern Argentina in South America, as the possible locations of common bean domestication; but this remains a matter of debate.

The debates go on, and breeders need to keep up to date with the scientific literature in order to understand which germplasm has been used during the evolutionary history of crop species, and the genetic resources not used to date, but which could be the key to future progress and success.

## ***Plants and People***

A broad overview of current ideas on how people and societies have evolved together with the crops upon which they depend can be found in the book *People, Plants & Genes* by Murphy (2007). A news focus on *Seeking Agriculture's Ancient*

*Roots* appeared in *Science* in June 2007 (Balter 2007). Interested readers can also locate written accounts of the agriculture of antiquity, such as the *Georgics*, the didactic poem of the Roman poet Virgil written around 30 BCE (Hahlbrock 2009). Mention is made in the poem of adding dung and ashes to arable soils and the use of legumes in crop rotation with cereals (King 2011). There will also be books of parochial interest, for me *The Scots: a Genetic Journey* by Moffat and Wilson (2011). *Feeding Frenzy* by McMahon (2013) opens with a brief history of food; explaining how farming communities and systems developed to select appropriate plants (and animals), manage water, renew soil fertility, protect crops from pests and diseases and to harness power to increase crop production. The history moves from fertile river valleys which provided water and nutrients; through the shifting of cultivation to new fertile soils by ‘slash and burn’; the use of livestock to restore fertility; the use of legumes that also restored fertility but also provided more animal feed; the mechanization of farming in the nineteenth century, accompanied by global transportation of farming products; to the modern agricultural revolution that dramatically increased yields per unit of land, at least in some parts of the world.

There is still scope for bringing more plants into cultivation as we have become dependent on relatively few of the 300,000 plant species accepted on “The Plant List” produced by a group of Botanic Gardens (<http://www.theplantlist.org>). The second edition of *Evolution of Crop Plants* by Smartt and Simmonds (1995), for example, considers about 230 crops to be of sufficient importance for inclusion, somewhat less than the 1000 studied by Vavilov for their origin and geography, and the estimated total of 2500 plant species that have undergone domestication worldwide (Meyer et al. 2012). Today a number of species are entering plant breeding programmes from previously undomesticated crops in West Africa (e.g. baobab and dika) and semi-domesticated ones in Amazonia. The scope for further domestication remains debatable because our ancestors appear to have known their plants and domesticated the obvious ones.

## Dispersion

### *Initial Spread of Farming*

Vavilov concluded that many species of cultivated plants remained within their geographical centres of origin. Those that were dispersed by humans, however, encountered different geographical areas and ecological conditions and became differentiated into geographical races and ecotypes, and then through further isolation into locally adapted landraces and strains. Thus the types of wheat and barley found in Abyssinia and China differ from those in their centres of origin in Asia Minor, as do those forms found in Europe. Wheat and barley also provide a good example of the rate of dispersion. As mentioned earlier, domestication took place in the ‘Fertile Crescent’ some 12,000–13,000 years ago, and from there, farming

spread east to India and China and west to Europe and North Africa. It had reached the Hungarian Plain some 7500 years ago (5500 BCE), and we know that emmer wheat, bread wheat and barley were being grown as crops in the east of Scotland 6000 years ago (Moffat and Wilson 2011). These grains were found in the postholes of a timber house which had been revealed to archaeologists by crop marks in a field at Balbridie, Aberdeenshire, and which had been dated to 3900 BCE. This was 4500 years after pioneer settlers arrived in Cramond, near Edinburgh, following the thaw of the tundra that had formed over Scotland around 9400 BCE. The tundra was a consequence of the massive volume of cold freshwater that poured into the Atlantic Ocean from the melting ice sheets over North America at the end of the last ice age.

Research by Sokal et al. (1991) provided support for the hypothesis that the spread of farming across the Middle East and Europe was associated with human population expansion, migration and intermating, and not simply caused by cultural diffusion (simple learning) of new techniques. Their evidence came from examining 26 polymorphic blood proteins of extant people from 3373 locations across the Near East and Europe. A specific example is provided by Moffat and Wilson (2011): they report that the male lineage (Y chromosome variant) of a farmer on the Hebridean island of Islay could be traced to immigrants whose variant chromosome originated in the Near East (modern Iraq) less than 10,000 years ago; in other words to the origin of farming in the ‘Fertile Crescent’ of the Tigris and Euphrates rivers of Mesopotamia. Interestingly, ryegrasses were probably spread as weeds of cereal crops by migrating farmers, but the earliest reference to the use of grass seed for the deliberate conversion of arable land into grassland was in 1677 in the UK (Humphreys et al. 2010). A further example comes from Skoglund et al. (2014) who managed to sequence the DNA from the bones of 11 early Scandinavian hunter-gatherers and farmers dating back to between 5000 and 7000 years ago. Four were associated with late Stone Age farming settlements and seven were identified as coastal hunter-gatherers. Genetic analysis revealed that the two groups descended from distinct genetic lineages and that there was gene-flow from hunter-gatherers to farmers but not vice versa. The authors interpreted their results as clear evidence that people from hunter-gatherer groups were incorporated into farming groups as they expanded across Europe. They also suggested that Stone-Age foraging groups were historically in low numbers, likely due to oscillating living conditions or restricted carrying-capacity due to food shortages.

### ***Exchanges Between New and Old Worlds***

Extensive dispersion and redistribution of crops followed the discovery of the New World by Columbus in 1492, often with political, economic and social consequences. Indeed, today many economically poor countries are among the richest in terms of the genetic diversity (resources) needed to ensure human survival (Esquinias-Alcázar et al. 2011). Five examples of dispersion will briefly be considered.

*Maize:* Following its domestication in Mexico, maize (*Zea mays*) spread widely throughout the Americas. Its dispersion throughout the world can be summarized from Tallury and Goodman (2001) as follows. Columbus saw maize in Cuba in 1492 and took it with him to Spain in 1493. By the end of the 1500s it was being widely grown in Spain, Italy and southern France. From the early 1500s onwards, Portuguese traders took maize from Spain and Italy to East, Central and West Africa and from the Guyanas and Brazil to West Africa. It was Dutch settlers who brought maize to South Africa, and imported it from the USA and northern Mexico around the turn of the twentieth century. Mediterranean traders probably introduced maize along the west coast of India in the early 1500s. It had been introduced into the Philippines by the time of Magellan's death there in 1521. By the mid-1600s it was well established in the Philippines, Indonesia and Thailand, and from there to China, Japan and Korea by the mid-1700s. Today maize is grown from 58°N to 40°S, in cool temperate to hot tropical and sub-tropical zones, and from tropical highlands to lowland regions. Globally, in 2012 it was the leading cereal grain crop. Production at 873 million tonnes was greater than rice (738) and wheat (671), and of the 873 million tonnes, 274 (31 %) came from the USA and 393 (45 %) from those countries now classified as least developed rather than developing (FAOSTAT 2012). Maize is mainly consumed as a staple food in Latin America, Africa and Asia, whereas in developed countries it is mostly fed to animals, but is also processed into many products including biofuel.

*Common bean:* Another of the first crops to be introduced from the New to the Old World was the common bean (*Phaseolus vulgaris*). The Mesoamerican common bean probably arrived in Europe through Spain and Portugal in 1506 and the Andean one in the same way after 1528, when the Spanish under Pizarro conquered Peru (Angioi et al. 2010). The pathways of dispersion across Europe were complex, with several introductions from America combined with direct exchanges between European and other Mediterranean countries. Adaptation to new environments and consumer preferences took place. Angioi et al. (2010) assessed the genetic structure and level of diversity of a large collection of European landraces of *P. vulgaris* (307) in comparison to 94 genotypes from the Americas that were representative of the Andean and Mesoamerican gene pools. They analyzed six chloroplast microsatellite (cpSSR) loci and two unlinked nuclear loci (for phaseolin types and *Pv-shatterproof1*). They were able to show that 67 % of the landraces were of Andean origin, and that there were no strong differences across European regions for the proportions of the Andean and Mesoamerican gene pools. Furthermore, the cytoplasmic bottleneck that was due to the introduction of *P. vulgaris* into the Old World was very weak or nearly absent. This was in contrast to evidence from nuclear analyses that had suggested a bottleneck of greater intensity. Finally, they estimated that 44 % of the European bean germplasm was derived from hybridization between the Andean and Mesoamerican gene pools. Although hybrids were present everywhere in Europe, they showed an uneven distribution, with high frequencies in central Europe and low frequencies in Spain and Italy.

*Sunflower:* Today sunflower (*Helianthus annuus*) oil is the fourth most important vegetable oil in world trade and is cultivated on over 22 million hectares, over half of which is in the Russian Federation, Ukraine, India and Argentina. The crop was domesticated in central-eastern USA, as already mentioned. Sunflower was introduced from North America into Europe by the early Spanish explorers in 1510, where it initially gained popularity as a garden ornamental. The agronomic development of sunflower as an oilseed crop took place in Russia, where a number of landraces had been produced by the late 1800s. Sunflower was reintroduced from Russia to North America in the latter part of the nineteenth century (Fernández-Martínez et al. 2009).

*Potato:* It is assumed that Pizarro and his men were the first Europeans to see potatoes (*Solanum tuberosum*) being cultivated, in the Andes of Peru in 1533 during their conquest of the Incas, but there is no written record (Hawkes 1990). Potatoes were actually first recorded in 1537 in what is now Colombia (Hawkes 1990). The first record of cultivated potatoes outside of South America was their export in 1567 from Gran Canaria in the Canary Islands to Antwerp in Belgium. This was 6 years before they were first recorded in Spain in 1573 in the market archives of the Hospital de La Sangre in Seville (Hawkes and Francisco-Ortega 1992). Potatoes were therefore probably first introduced from South America into the Canary Islands around 1562 and from there to mainland Europe (Hawkes and Francisco-Ortega 1993). It has often been assumed that these early introductions came as ships' stores from Colombia and were of Colombian or possibly Peruvian origin, and hence were primarily tetraploid group Andigena potatoes. Then as the growing of potatoes spread north-eastward across Europe, they became adapted to the long summer days of northern Europe and in this respect resembled Chilean potatoes. However, extant Canary Island potatoes comprise both Andean- and Chilean-type landraces and Rios et al. (2007) have suggested that there were multiple early introductions of both types. Furthermore, they suggested that the early European potato was selected from the Chilean introductions because they were better adapted to European conditions. It now seems safest to assume that the early introductions of cultivated potatoes to Europe came from both the Andes and coastal Chile. Analysis of DNA from 49 herbarium specimens has confirmed the presence in Europe of Andean potatoes from around 1700 and Chilean potatoes from 1811 (Ames and Spooner 2008). After their introduction to Europe, potatoes initially remained a botanic curiosity, being grown and studied in physic gardens for interest and medicinal purposes. It was food shortages during the eighteenth century that proved to be the stimulus to potato cultivation throughout Europe because military and economic strength depended upon adequately fed manpower. During the seventeenth and eighteenth centuries many European countries developed widespread political and commercial interests in the rest of the world, and the European colonists and missionaries took with them their common crops including potatoes. Potato production expanded worldwide during the nineteenth century, but it was the expansion in China and India during the second half of the twentieth century that led to these countries becoming the two most important producers in

the world (Bradshaw and Bonierbale 2010). Cassava (*Manihot esculenta*) also became an important staple food in the tropics after it was taken from Latin America to Africa and then Asia.

**Wheat:** The nineteenth century saw the spread of European settlers with their crops across the globe, including the Great Plains of the USA. Here grassland and savannah were transformed in a relatively short period of time into a wheat (*Triticum aestivum*) growing region that has become one of the most productive breadbaskets in the world (Kingsbury 2009). Likewise during the nineteenth century, wheat growing became established in Australia, now a major exporter.

### ***Colonialism and Tropical Plantation Crops***

Today tropical plantation crops are often grown far from where they were domesticated as a result of the activities of European colonialists; as will be seen in the following examples. Some of these crops are associated with the darker side of human history, as well as with some important historical events.

**Sugar cane and tobacco:** The year 1612 saw the start of tobacco (*Nicotiana tabacum*) growing in what was then the British colony of Virginia whilst in 1642 the long and complex journey of sugar cane (*Saccharum officinarum*) growing reached Barbados. Subsequently very profitable trades to Britain developed in tobacco and sugar as commodities; the dark side of which was a human slave trade, made possible by British and other European colonial exploits in Africa, and which was first ended in the British Empire by Act of Parliament in 1807. The 200-year shipping of Africans across the Atlantic to supply the sugar industry of the Caribbean and Americas with labour created a massive diaspora whose effects are still felt by many millions of people today (Benson 2012).

**Tea and coffee:** Tea (*Camellia sinensis*) drinking had flourished in China since the Tang (618–906 CE) and Song (960–1279 CE) dynasties, having originated much earlier in the southwest of the country. It increased in popularity in Europe and North America from the seventeenth century onwards, with China the sole supplier. The British East India Company eventually established a monopoly on the tea trade, and surreptitiously raised the money to buy the tea through illegal opium imports into China that were destabilizing its population through opium addiction. It was the infamous Opium Wars with Britain from 1839 to 1842 that forced China to open up trade; but by then the monopoly of the British East India Company had ended. Subsequently, Robert Fortune established the East India Company's tea industry in northern India (Assam and Sikkim) in the second half of the nineteenth century from 23,892 young plants and approximately 17,000 seedlings he had gathered in China in 1848 (Musgrave et al. 1998). At about the same time a type of tea plant native to Assam was discovered and 'seed gardens' were established with pure China, pure Assam and mixed plantings from which Indian hybrid tea was formed

(Ellis 1995). Likewise in the 1870s, tea plants collected in China and cultivated in the Botanic Garden in Ceylon (now Sri Lanka) were used to establish a tea industry in that country. By 1900, 40 % of the tea consumed in Britain came from Ceylon (Parry 2004). Hybrid tea seed from India was also taken to Ceylon (Ellis 1995). However, perhaps the biggest effect of tea on the course of human history was the British tax of 1773 on tea exported to its American Colonies which contributed to the American Declaration of Independence on the 4th of July 1776. As a consequence it is said that Americans prefer coffee to tea.

Interestingly, coffee provides a recent example of a crop intimately tied up with politics, economics and environmental issues. Today Vietnam is second in importance to Brazil as an exporter of coffee, albeit focused on poor-quality robusta (*Coffea canephora*) beans for instant coffee, rather than arabica coffee (*C. arabica*). The crop was introduced to Vietnam in 1857 by the French and slowly grew in importance. The Vietnam War which ended in 1975 disrupted production in the Buôn Ma Thuôt plateau region where the industry was centred. Collectivization of farming by the Vietnamese after 1975 did not help production, whereas from 1986, privately owned enterprise was permitted by the Communist Party and saw a surge of growth in the industry. Over reliance of a country on a single crop for export is a cause for concern, and so are the environmental issues associated with a large scale plantation crop, but the story is intriguing (Chris Summers, BBC News 25 January 2014).

*African oil palm and rubber:* The centres of origin and diversity of African oil palm (*Elaeis guineensis*) are concentrated in the tropical forests of West and Central Africa, in forest clearings made by human disturbance. It was European colonists who started oil palm plantations in Indonesia and Malaysia to ensure a steady supply of oil as an industrial crop during the later part of the nineteenth century (Hardon 1995; Soh et al. 2009). Remarkably, the industry was based on hybridizations and selections from four seedlings planted in Bogor Botanic Gardens in Indonesia, apparently derived from the same fruit bunch in West Africa, and obtained via Amsterdam and Mauritius/Reunion in 1848. These were distributed to plantations in Deli province in Sumatra and from there to Malaysia, and became the commercial planting material for the rapidly developing plantation industry in Malaysia and Indonesia from 1911 to the early 1960s. These two countries still have the bulk of the global commercial plantings, now composed of thin-shelled Tenera hybrids from crossing the original thick-shelled Dura fruit form with the shell-less Pisifera type. Another plantation crop that reached Malaysia via botanic gardens was rubber (*Hevea brasiliensis*) (Parry 2004). In the mid-1870s, botanists Joseph Hooker and Clement Markham devised a scheme to collect seeds in the plant's native Brazil and to germinate them at the Royal Botanic Gardens at Kew. Seedlings were then transferred to botanic gardens in Sri Lanka and Singapore, and from there to plantations in Malaysia where they were cultivated en masse. Today Thailand, Indonesia and Malaysia are the main producers of natural rubber in the world.

## Adaptation

Plant breeding results in the adaptation of cultivated plants to new environments and new human uses; one could say to whole new farming and production systems. Three phases can be recognized in its history. The first one lasted until the beginning of the nineteenth century and was characterized by the practice of selection on the extensive, naturally occurring genetic variation within cultivated species. For most of this period plant breeding was done by farmers; indeed one could almost define this phase as farmer selection resulting in numerous landraces. They practised both unconscious and deliberate selection when they kept some of the harvest for planting or sowing their next crop. In addition, natural selection would be operating on the genetic diversity of the crop in new environments, during both domestication and subsequent dispersion. Some genetic variants were better able than others to survive to harvest, and from them, some were more productive than others. Selection remained an important aspect of the second and third phases of plant breeding in which planned hybridization was added, followed by scientific breeding methods based on genetic knowledge. Indeed, whatever the strategic goals of any breeding programme, they have to be translated into selection criteria that the breeder can actually use, and a prerequisite is often greater scientific understanding followed by the development of simple rapid screens of breeding material.

## Selection During Domestication

It is perhaps rather academic to debate the relative contributions of natural, unconscious and deliberate selection to crop improvement, particularly for traits associated with sowing and harvesting crops. The transformations of many plant species during domestication are clear and have been extensively discussed in the scientific literature and books on the subject, such as those by Murphy (2007), Kingsbury (2009) and Hancock (2012). Furthermore, the genetic bases of these transformations are being discovered and understood, with the potential for additional systematic manipulation in future crop improvement (Doebley 2006), as we will see in Chap. 5. Two types of examples will suffice at present, one from cereal domestication and the other from potato domestication.

### Cereal Domestication

Domestication of rice (*Oryza sativa*) has resulted in alterations to a large array of morphological traits, including loss of seed shattering, change in grain coloration from black to straw-white, enlargement of seed size and change from prostrate to erect growth habit (Zhu et al. 2011), as well as reduced seed dormancy (Vaughan

et al. 2008). The seed-shattering habit is considered a key functional trait for wild rice (and other cereals) to survive efficiently and the black hull is thought to be a natural colour for shattered seeds in a dark mud land. In contrast, the non shattering seed of cultivated rice, essential for harvesting, is always associated with a straw-white hull, which is accompanied by a withering of stems and leaves of a similar colour at the end of the growing season. In Chap. 5 we shall see that seed shattering in wild rice is controlled by the *Shattering4* (*Sh4*) locus on chromosome 4 and black hull by the separate *Black hull4* (*Bh4*) locus also on chromosome 4. Even bigger differences between a cultivated crop and its wild ancestors are those between maize and teosinte (Hancock 2012), and it was native farmers in America who selected the types (races) of corn that we recognize today: flint, flour, gourdseed dent, pop and sweet corn.

## **Potato Domestication**

A second type of change during domestication concerns the consumption of the crop, namely the loss of toxic compounds in the parts of the plants which are eaten. The tubers of many of the wild relatives of our cultivated potato contain potentially toxic levels of glycoalkaloids. It has therefore been assumed that domestication by Andean farmers involved selection for less bitter and hence less toxic tubers. Interestingly, Johns and Alonso (1990) found that some genebank accessions of *S. bukasovii* had tuber glycoalkaloid levels which were consistently close to those found in many cultivated (diploid) potatoes. Hence domestication of this species would have required little or no selection for lower glycoalkaloid level, unlike *S. canasense*, *S. leptophyes* and *S. sparsipilum*, which have been suggested as possible progenitors. So domestication of potatoes may have involved greater farmer selection between wild species than within them.

## **Visual Selection**

### **Potatoes, Tomatoes and Carrots**

*Potato:* Potato tubers also provide a good example of deliberate visual selection. Potato tuber diversity was in fact one of many examples given by Darwin, in *The Variation of Animals and Plants under Domestication* (Darwin 1868), of the principle that the valuable and selected parts of all cultivated plants show the greatest amount of modification. The wild relatives of cultivated potatoes have small tubers, round or oval in shape, with white flesh and skins that are white or have a faint bluish-purple flush. In contrast, Andean farmers retained a much wider variety of tuber shapes and skin and flesh colours (Glendinning 1983; Simmonds



**Fig. 1.2** Tomatoes on sale in Rovigo, Italy (2 October 2014)

1995a), and this diversity must have arisen by genetic mutation during domestication. The diversity was greater than we see today in modern cultivars. Tuber shapes were either regular, from compressed through elliptic to long ones, or unusual irregular ones such as coiled and concertina-shaped. Skin colour was pink, red, blue or purple, due to anthocyanins, or white, and the distribution of pigments could result in pigmented eyes or eyebrows, and splashed, scattered, spectacled, or stippled tubers. Flesh colour was yellow or orange, due to certain carotenoids, or white, and anthocyanin pigmentation could also occur (Ortiz and Huaman 1994).

**Tomato:** We can also see the effects of selection in a close relative of the potato, namely the tomato (*Solanum lycopersicum*). In contrast to its wild relatives that have round, two-loculed fruits, the cultivated tomato fruit (Fig. 1.2) is highly diverse in shape and colour as well as having increased weight (Rodríguez et al. 2011). Today, for example, tomatoes can be classified into eight categories for fruit shape: flat, round, rectangular, ellipsoid, heart, long, obovoid, and oxheart. Later we shall see that about 70 % of the variation in fruit shape found in a diverse collection of 368 tomato accessions could be explained by combinations of variants that arose in just four genes (Rodríguez et al. 2011). More recently, Visa et al. (2014) analysed fruit shape using elliptic Fourier shape modelling and Bayesian classification of contour morphometric data, and identified long rectangular as an additional ninth shape. Morphometric analysis uses the position of, and distance between, landmarks of an object as the source of morphological data. The researchers analysed 784 longitudinal sections from 48 phenotypically distinct

tomato cultivars grown in three locations. They found that the flat and ovoid categories contained fruit that consistently conformed exclusively to a single category. In contrast, the categories oxheart and long rectangular featured fruit that tended to equivalently fit several categories of shape. Classification of tomato plants based on fruit shape is critically important for the correct categorization of different cultivars, both for the certification of new cultivars and for determining their method of harvest (machine or hand-picked) and end use (processed into paste and sauce, canned and diced, eaten fresh or sliced for sandwiches and hamburgers).

*Carrot:* An interesting change in human colour preference can be seen in the carrot (*Daucus carota*). Wild carrot roots are white or very pale yellow, whereas purple and yellow were the colours of the first domesticated carrots, with purple more popular east of Central Asia and yellow more popular in the west (Banga 1957a). Orange carrots were not described until the sixteenth and seventeenth centuries, but were soon preferred in both eastern and western production areas (Banga 1957b).

## Selection in New Environments

Many examples of selection in new environments can be found in the dispersion and redistribution of crops following the discovery of the New World. The pre-Columbian spread of maize, from its centre of origin in tropical southern Mexico to the higher latitudes of the Americas, required post domestication selection for adaptation to longer day lengths. We shall examine the genetical basis of this in Chap. 5. The potato was introduced from South America to Europe during the latter part of the sixteenth century, and from there to the rest of the world from the seventeenth century onwards. First in South America and then worldwide, locally adapted potatoes were selected that could produce tubers in different day lengths so that today potatoes are grown worldwide from latitudes 65°N to 50°S. In Chap. 5 we shall see that maturity in potato is controlled by the *StCDF1* locus on chromosome 5, and selection of genetic variant *StCDF1.2* at this locus results in early maturity and the ability to tuber in long days (Kloosterman et al. 2013). There was also selection for different growth cycles with lengths that matched the available growing seasons. These range from 180 days in summer in the high Andes, through 150–120 days in summer in the lowland temperate regions of the world, to 120–90 days in winter in the lowland subtropics where they can be as short as 75 days (Bradshaw and Bonierbale 2010). Likewise, the tomato was cultivated and consumed in Mexico long before the arrival of the Spanish. Its introduction to Europe and Asia were accompanied by further selection which resulted in different types associated with new uses and growing systems (Díez and Nuez 2008).

Wheat provides another example. Adaptation to the Great Plains of the USA required selection of cultivars that could cope with the droughts, early frosts and

shorter growing seasons of the plains, as well as rust epidemics. Likewise, Canada needed cultivars that could cope with a shorter growing season than any in Europe (Kingsbury 2009). Finally, in Brazil since the 1970s, the vast central plains of grassland and scrub, known as the Cerrado, have been transformed into large, mechanized farms growing soybeans, maize, beans and rice (McMahon 2013). This was primarily made possible by adding lime and phosphorus to improve the soil, but also involved development of tropical versions of soybean (*Glycine max*), a plant that originated in temperate China.

## Selection for Use in New Farming and Production Systems

Selection for use in new farming and production systems can be seen with fodder for livestock in Europe from the sixteenth century. The most fertile soils were used for arable crops with 2 years of cereal crops followed by 1 year of fallow. Livestock grazed in forests and on less fertile semi-natural grassland managed as common land, and was rested on the fallow plots overnight, fertilizing them with their excrements. As soon as it became possible to harvest and transport hay, livestock could be kept in stables during winter. As a result, more manure was distributed on the arable land, but grasslands became depleted in nutrients and this restricted stocking densities. At the end of the sixteenth century, initially in Flanders, farmers started to grow fodder crops (e.g. turnips) as catch crops after the cereal harvest and to grow nitrogen-fixing forage crops (e.g. red clover) on the fallow land. Both the catch crops and the red clover allowed livestock numbers to be increased, with a resultant increase in manure production. Furthermore, the clover left nitrogen in the soil for the subsequent cereal crops. These developments led to a doubling of cereal yields from an average of 1 t/ha up to 2 t/ha (Reheul et al. 2010). Later in Britain during the eighteenth century (Riddet 1925), Charles Townshend and Thomas Coke, both of Norfolk in England, convinced British farmers of the value of a four-course rotation of wheat, turnips, barley and clover. The turnips were drilled in rows wide enough apart for horse-hoeing of weeds, both drill and horse-hoe being early eighteenth century innovations of Jethro Tull. The turnips were either grazed *in situ* by folded sheep or lifted to fatten cattle in winter fold yards, something that had not previously been possible. Thus the turnip provided a break crop between cereals, a cleaning crop and good fodder for wintering livestock. These improvements in agriculture provided increased quantities of food for an expanding European population during the nineteenth century. Then the twentieth century saw the advent of new farming methods that used chemical fertilizers and selective herbicides. These changes removed the cultural need in arable systems for the traditional break crop, which could therefore be replaced by a cash crop such as sugar beet. Nevertheless, by 1910 in the UK and elsewhere, the grass-clover crop (ley) had become well established within a rotational system where 2–3 year old leys were ploughed up to provide sufficient nutrients for two or even three good cereal crops of wheat,

oats or barley (Humphreys et al. 2010). The grass-clover leys have continued to provide cheap feed for livestock and the promotion of soil fertility, and still have an important role in sustainable animal production.

Given that these new farming and production systems developed in Europe from the sixteenth century onwards, the achievements of the Incas in the Andes during the fifteenth century are impressive (Moseley 2001). They developed a sophisticated agriculture appropriate for the high altitude, cold and dry climate around their Peruvian capital of Cuzco. They were able to grow crops of maize, potatoes and other food plants during summer on terraces built of dressed stone which trapped the heat of the sun and aided irrigation; then store the produce for use during winter. The Incas also had a suite of locally domesticated plants and animals. Perhaps there are lessons for us today when thinking about what farming systems are appropriate for different parts of the world.

Selection for use in new farming and production systems continues today. A good example is breeding rice for short growth duration (Virmani and Ilyas-Ahmed 2007). Before the Green Revolution of the 1960s and 1970s, cultivars in tropical and sub-tropical Asia matured in 160–170 days and many were photoperiod-sensitive. This long duration meant that multiple cropping was not possible. The first cultivars of the Green Revolution matured in about 130 days with subsequent ones in 115–130 days; but even these could be grown only once a year. Today short duration cultivars are available that mature in 105–110 days and this has led to major changes in cropping patterns in Asia. An upland crop can be grown before or after the main rain fed crop, two crops can be grown in a rainy season, and in irrigated areas three crops can be grown in 1 year. Virmani and Ilyas-Ahmed (2007) reported that short-duration cultivars were very popular in Indonesia and Vietnam.

## Selection for Altered Chemical Composition

Good examples of deliberate human selection can be found in the planned alteration of the chemical composition of plants, for example in sugar beet and oilseed crops.

*Sugar beet:* After the discovery that fodder beet (*Beta vulgaris*) contained the same kind of sugar (sucrose) as sugar cane, sugar beet cultivars were selectively bred in Germany toward the end of the eighteenth century (Biancardi et al. 2010). The first beet sugar factory opened at Cunern (Silesia) in 1802. The rapid development of the sugar beet crop on mainland Europe was initially driven by a decline in access to cane sugar, the result of a blockade of French shipping ports during the Napoleonic Wars (1803–1815) between France and Britain (Benson 2012). New analytical and breeding methods were developed in France and sugar content was dramatically increased from 7 to 16 % during the late nineteenth century (Robb and Wishart 1915). This was achieved by repeated selection of plants with a high content as seed parents of the

next generation. The importance of sugar beet to Europe increased again during the First World War when the supply of cane sugar was threatened once more. Today sugar from both cane and beet is the world's predominant sweetener, with over 130 million tons consumed globally each year (Benson 2012). In 2002 the World Health Organization recommended that sugar should account for no more than 10 % of the calories in the human diet because of concerns about obesity, type-2 diabetes and damage to teeth; and in 2014 there was talk of lowering the guidance to 5 %.

*Oilseed crops:* Moving into the second half of twentieth century, oilseed rape (*Brassica napus*) provides another good example of selection for altered chemical composition. Over the course of the last 30 years, starting in Canada, this temperate crop was transformed into the world's third most important source of both vegetable oil and animal meal. In the 1960s and 1970s government researchers in Canada bred a new type of rapeseed, canola, whose products were low erucic acid oil (<1 % of fatty acids compared with original 45 %) and low glucosinolates meal (<18  $\mu\text{mol g}^{-1}$  seed at 8.5 % moisture), thus making them suitable for use as edible and cooking oils, and animal meal, respectively. Erucic acid, although a monounsaturated fatty acid, has a bitter taste and was thought to have a possible damaging effect on the heart muscle of animals. In monogastric animals the digestion of glucosinolates results in the release of toxic by-products that can cause liver and kidney damage along with lymph dysfunction (Friedt and Snowdon 2009). Again new analytical and breeding methods were the keys to success (Khachatourians et al. 2001). Conventional canola oil now comprises 60 % oleic acid (monounsaturated), 21 % linoleic acid (polyunsaturated), 10 % linolenic acid (polyunsaturated), 4 % palmitic acid (saturated), 2 % stearic acid (saturated) and 3 % other fatty acids (Friedt and Snowdon 2009). We want to minimize saturated fats in our diets because they are associated with heart disease through raised blood-cholesterol levels. In contrast, both linoleic and linolenic acid are essential fatty acids in our human diet. Despite the beneficial nutritional properties of linolenic acid, the oxidative stability of the oil can be improved by decreasing the linolenic acid content to less than 3 %, which results in enhanced shelf life and a reduction of *trans*-fatty acids (as opposed to the *cis* isomer) which are thought to increase the risk of coronary heart disease. The latter is a particularly important nutritional quality characteristic for high-temperature frying oils in the fast-food and food-processing industries. Today high oleic acid (>75 %), low linolenic acid (<3 %) cultivars (HOLL types) provide an important new quality of rapeseed oil for nutritional purposes (Friedt and Snowdon 2009; Despeghel and Guguen 2012).

Sunflower (*Helianthus annuus*) was also transformed into a major oilseed crop in the second half of the twentieth century due to a dramatic increase of oil percentage and the development of hybrid cultivars (Fernández-Martínez et al. 2009). The seed oil of traditional sunflower cultivars has a linoleic acid content of between 45 and 70 % and is considered a healthy vegetable oil suitable for salad and margarine production. Today mid- and high-oleic cultivars are available with oleic acid contents of 55–75 % and 85–90 %, respectively. These oils have a better thermooxidative stability, which makes them more appropriate for frying purposes.

Later, we shall also see that altering the chemical composition of the maize kernel for high and low contents of both oil and protein is the longest running selection experiment to date in plants, having continued for over 100 generations since 1896 (Dudley and Lambert 2004).

## Plant Breeders and Commodity Traders

In selecting for altered chemical composition, the activity of plant breeding has become divorced from the activity of farming. This is an example of a more widespread transfer of plant breeding from farmers to specialist breeders, at least in Europe and North America. Details about the history of these developments can be found in the book *Hybrid* by Kingsbury (2009). Briefly, the process began in Europe during the seventeenth and eighteenth centuries with the development of a trade in seed of cereal, fodder and vegetable crops. The same period also saw the start of a transition to farmers and growers producing crops to sell, particularly to expanding urban populations. As a result, the numbers employed in the agricultural sector in Central Europe fell from 80 % of the population in 1800 to less than 5 % today (Hahlbrock 2009). Thus seedsmen grew crops for the production of quality seed which was bought by merchants and sold to farmers and growers, who in turn produced crops to sell to the consumer.

Today trade in food is international and until recently dominated by a small number of multinational corporations: Cargill, Archer Daniels Midland, Louis Dreyfus, Bunge and Glencore; in other words, just five commodity trading companies (McMahon 2013). The oldest, Bunge, was founded in Amsterdam in 1818. Thus the global food system now resembles an hourglass: hundreds of millions of farmers growing food which is mainly supplied to hundreds of millions of consumers through just a few commodity traders. It is thought that the five firms just mentioned handle more than four-fifths of the world's total trade in grains and oilseeds (McMahon 2013). Despite concerns about their dominance, these companies play an important role in facilitating the physical trade in food: buying, processing, moving and selling food across borders. As long as some countries have food deficits and others food surpluses, there will be a need for intermediaries to facilitate trade. Now, however, the five companies do face competition from new Asian companies such as Noble, Olam and Wilmar, as well as ones in China, Japan and South Korea. These new Asian food traders want to secure full control (ownership) of the supply chains, from land use to the end user; in other words, vertical integration including owning and operating farmland. The older companies are following suit (McMahon 2013).

We must now return to our main story. A steady stream of seed companies were established between the middle of the eighteenth century and the end of the nineteenth century, for example: Vilmorin in France (founded 1743) and Gartons (1898) in Britain, for cereals; and Suttons (1806) in Britain and Burpee (1876) in the USA, for vegetables. Trade in seed led to the naming of cultivars and by the

late-nineteenth century an enormous number were being promoted in seed catalogues, having been produced by farmers and growers, hobby breeders, seedsmen, and seed companies. However, some were chance discoveries. This was true of the three cultivars of wheat most extensively cultivated in Scotland in the middle of the nineteenth century. Hunter's wheat was raised from a single plant found by the roadside, Fenton from three ears which grew in an old quarry, and Hopetoun from a single plant found near a village close to Edinburgh. The plants were observed to possess the characteristics most approved by farmers and were simply used as the stock from which a new cultivar was propagated (Anderson 1855). Interestingly, there is documentation of such single plant selection as early as 1662–1722 in China (Binglun 2009); a reminder of how advanced Chinese agriculture was at that time. The nineteenth century in Europe and North America did however see an important development in the scientific improvement of cultivated plants, planned hybridizations, whether natural or artificial. By then there had already been over 100 years of progress in understanding sexual reproduction in plants.

## Sexual Reproduction in Flowering Plants

The prerequisite for planned hybridizations in plant breeding was an understanding of sexual reproduction in flowering plants. In his book *The Anatomy of Plants* (1682), the English botanist, physician and microscopist Nehemiah Grew (1641–1712) correctly identified the stamens and pistils as the male and female sex organs in flowers. Stamens are known collectively as the androecium and carpels rather than pistils as the gynoecium. Rudolph Camerarius, in 1694 in Tübingen, described their function in fertilization and showed that pollen was required for the process (Carlson 2004). Complete flowers have sepals and petals in addition to stamens and pistil, whereas incomplete flowers lack one or more of these floral organs. For example, crops belonging to the grass family lack petals and sepals, and so do buckwheat and sugar beet. The floral organs are attached in whorls to a receptacle at the tip of a flowering shoot. Working from the outside to the centre we have the following whorls (Fig. 1.3). Sepals (collectively, the calyx) enclose the flower in bud and are usually small green leaf-like structures below the petals. Petals (collectively, the corolla) are the conspicuous portion of most flowers, often being highly coloured, and may contain perfume as well as nectar glands. Each stamen has a pollen-producing anther on a stalk called a filament. Finally each pistil has a pollen-receptive area, the stigma, borne on a stalk called a style that extends from the ovary. The ovary has one or more ovule-bearing units called carpels with the ovules attached via a placenta (Moore et al. 1998). The reader does, however, need to be aware of the fact that botanical terminology can vary slightly. *The Cambridge Illustrated Glossary of Botanical Terms* (Hickey and King 2000) defines the carpel as one of the units forming the gynoecium, usually consisting of ovary, style, and stigma. The pistil is then defined as a single carpel in an apocarpous (having free carpels) flower, or the gynoecium in a syncarpous (having



**Fig. 1.3** Oilseed rape (*Brassica napus*) flower showing four yellow petals (above four hidden green sepals) in the shape of a cross, with a central style and stigma surrounded by six stamens (four long and two short)

united carpels) flower. For the purpose of this book, we need to focus on the stigma, style and ovule-containing ovary, and can simply refer to the overall structure as a pistil.

Pollination was studied in detail by the Swedish botanist Carl Linnaeus (1707–1778) and the German botanist Joseph Kölreuter (1733–1806). However, the Italian microscopist Giovanni Amici was the first to clearly identify a pollen tube in 1823, and in 1847 to observe its entry into the ovule, and into the egg-containing embryo sac (Carlson 2004). Kölreuter also realized that pollination was not limited to the distribution of pollen by the wind; he recognized the role of insects in bringing pollen to flowers (Carlson 2004). After pollination the ovules develop into seeds and the ovary becomes the fruit.

Linnaeus also observed that flowering plants could be divided into hermaphrodite, monoecious, dioecious and polygamous species. The vast majority (about 90 %) of flowering plants are hermaphrodites (Charlesworth and Charlesworth 2010); all of the flowers on a plant have both male and female sex organs and self pollination within a flower is a possibility. Some flowering plants like maize have separate male and female flowers on the same individual (monoecious), as recognized by Camerarius. Others have separate male and female individuals (dioecious): for example, hemp was studied by Camerarius and date palm by ancient Assyrian and Medieval Arab writers who recognized separate fruit and pollen producing plants (Kingsbury 2009). Other dioecious species are papaya, spinach and asparagus. Polygamous species have male, female and hermaphrodite flowers on the same or on different plants. Botanists, however, prefer to talk about staminate and pistillate flowers and plants rather than males and females.

Descriptions and photographs of the great diversity of flower structure can be found in botany textbooks (Moore et al. 1998).

## Classification and Naming of Plants

Linnaeus used variation in flower structure to group and hence to classify plants; in total 7700 species. He published the first edition of his *Systema Naturae* in 1735 and his two volume work *Species Plantarum* in 1753. His classification was too simplistic, being based on a few key features of flowering plants. In contrast, the English naturalist John Ray had grouped plants on the basis of multiple similarities, and had given a detailed classification of more than 18,000 kinds of plants in his three volume *Historia Plantarum Generalis* which was published in 1686. However, it was Linnaeus who persuaded botanists to adopt a standardized nomenclature for plants in which each species was given a Latin name consisting of two words, the genus and species.

Later, in the early nineteenth century, the successors of Linnaeus replaced his artificial classification of plants with a more natural one, similar to that of Ray. Consideration of the similarities and differences among individuals and groups of plants leads to a hierarchy of categories called ranks. At each level a number of smaller groups are brought together to make a large one. Species are at the lowest level and domains (Eubacteria, Archaea and Eukaryota) at the highest level and in between we have genus, subtribe, tribe, subfamily, family, order, class, division and kingdom; each category is a taxon. Within the Eukaryota are six kingdoms of which the Archaeplastida includes red algae, green algae, freshwater algae and all land plants (Blackmore 2012). The Archaeplastida all have chloroplasts derived directly from blue-green algae (cyanobacteria), resulting in the presence of a double membrane around each chloroplast. The Viridiplantae (green algae and all land plants) or green plants all have chloroplasts with double membranes and both chlorophyll a and b. Finally, all land plants are embryophytes because their embryos grow within the archegonia, the structures containing the egg cells.

After Darwin published *On the Origin of Species By Means of Natural Selection* in 1859 (Darwin 1859), systematists developed the natural classification of organisms in terms of evolutionary relationships and history, resulting in the pedigree or genealogical history that we call the “Tree of Life”. One phylogenetic approach to classifying organisms, named cladistics, results in clades defined as any set of organisms with a common ancestor. Since 2000, there has been heated discussion on whether or not the phylogenetic classification should be accompanied by a PhyloCode naming system to replace that of Linnaeus (Pennisi 2001). The third edition of the *Oxford Dictionary of Plant Sciences* (Allaby 2012) does include an appendix with the Plant Classification system known as Angiosperm Phylogeny Group III, which was published in 2009, and is continually updated on its website (<http://www.mobot.org/MOBOT/research/APweb/>). It treats the major monophyletic groups of land plants as ‘subclasses’ with the flowering plants as subclass

Magnoliidae. The 16 major branches within the flowering plants are treated as superorders. Starting at the base of the ‘tree’ the first seven superorders are: Amborellanae, Nymphaeanae, Austrobaileyanae, Magnolianae, Liliinae (corresponds to the monocotyledons), Rosanae and Asteranae. The dicotyledons are not a monophyletic group and hence are divided between lineages with most belonging to the large and very diverse superorders Rosanae and Asteranae (Blackmore 2012). A thorough account and discussion of plant nomenclature and taxonomy can be found in the review by Spooner et al. (2003). In this book I have stuck with Linnaeus and where appropriate given the Latin names (*Genus species*) of cultivated plants in brackets after their common English names. References are also made to families (end with ‘-aceae’).

## Species Concepts

The period from the mid eighteenth to the mid nineteenth centuries saw extensive experimentation in plant hybridization, first by naturalists and then also by breeders. Accounts of the large number of inter-specific hybridizations done by Linnaeus and the German botanists Joseph Kölreuter and Karl Friedrich von Gärtnér can be found in the books by Carlson (2004) and Kingsbury (2009). The experiments revealed that hybridization was only possible between closely related species and that many vigorous hybrids were sterile. A famous early example was the sterile hybrid pink of Thomas Fairchild (1667–1729), a cross between a Sweet William (*Dianthus barbatus*) and a carnation (*D. caryophyllus*), which was reported at a meeting of the Royal Society in London in February 1720 (Kingsbury 2009). In summary, observations of reproduction led to the biological species concept of distinct reproductive groups with barriers to hybridization between them. When these groups could be distinguished by prominent morphological discontinuities, taxonomic classification recognized the same species, but the two do not always coincide. It was found that taxonomic species belonging to distinct genera could only rarely be crossed, and those belonging to distinct families could never be crossed. Phylogenetic approaches to classification have also led to a third and distinctive evolutionary species concept (Grant 1971) which embraces a greater diversity of breeding systems than the original concept of biological species, which applied to biparental organisms.

When considering genetic resources, plants breeders are interested in the crossability of cultivated plants with their wild relatives and trying to overcome barriers to hybridization. We shall see later that they find the concepts of primary, secondary and tertiary gene-pools useful, but need to report their results in terms of recognized species names which are currently based on the taxonomic system of Linnaeus. The potato and its wild relatives provide a good example of the species concepts and all of these issues, as explained in the review by Camadro et al. (2012).

## Deliberate Hybridization in Plant Breeding

Deliberate hybridization in plant breeding started at the beginning of the nineteenth century and was common practice by its end. The aim was to make a cross between two parents and then to select from the offspring, a new cultivar combining the desirable characteristics of each parent but lacking the undesirable traits. One of the first British hybridizers was a country gentleman, Thomas Andrew Knight, who worked with fruit and vegetables, such as apples and garden peas, and was president of what became the Royal Horticultural Society (RHS) from 1811 to 1838. He advocated the use of artificial hybridization in the breeding of new potato cultivars, but James Clark in England was probably the first breeder to produce successful cultivars from such deliberate cross-pollinations, for example, Magnum Bonum in 1876. Many nineteenth century maize farmers made deliberate crosses between cultivars, but it was William Beal (1833–1910) of Michigan Agricultural College who first published results in 1876. He crossed maize by growing two cultivars together in isolation from others, and then removing the male flowers (detasseling) of every plant of one cultivar so that its female flowers could only be fertilized by pollen from the other cultivar. Patrick Shirreff of Hopetoun, Scotland, used hybridization to produce the wheat cultivar Shirreff's Squarehead in 1882, and in 1883 the Vilmorin family in France produced the first modern French wheat cultivar Dattel from a cross between French and British wheat. These are just a few of the many examples to be found in Kingsbury's book *Hybrid* (Kingsbury 2009).

It was the RHS that organized the First International Conference on Hybridization and Cross-Breeding in London in 1899. By then the first books had been written on plant breeding: *Instructions for cereal breeding on a scientific and practical basis*, in German by Kurt von Rümker at Göttingen University in 1888, and *Plant-breeding, being five lectures upon the amelioration of domestic plants*, in English by Liberty Hyde Bailey at Michigan State Agricultural College in 1894 (Kingsbury 2009). One of the first university courses in plant breeding was established by Rollins Adams Emerson at the University of Nebraska in 1899 (Kingsbury 2009). Plant breeding was thus established both as a discipline and as a profession.

### ***Mechanics of Artificial Hybridization***

Today the breeder of any particular plant species will have a detailed knowledge of its floral anatomy and physiology and the mechanics of artificial hybridization. The breeder will know about any vernalization and day length requirements to induce flowering; when ripe pollen is shed from the anthers and when the stigma is receptive to pollen; how to store pollen when the two parents flower at different times; the need for accurate labelling of pollinated flowers; the need to collect seed before it is naturally shed; and finally how to extract and store seed. Of particular

importance are the precautions required to prevent unwanted pollinations, such as the exclusion of insects, the emasculation of hermaphrodite flowers before they open, covering flowers with bags, and cleaning brushes in alcohol before collecting pollen from a different source. The interested reader can get a better idea of exactly what is involved from books on practical breeding such as *Hybridization of Crop Plants* (Fehr and Hadley 1980) and *Breeding Ornamental Plants* (Callaway and Callaway 2000), which covers 15 groups of plants and for each gives a detailed description of hybridization mechanics.

## Mendel and the Laws of Inheritance

All of the nineteenth century experiments on plant hybridization, including those done by Knight and Darwin on the garden pea (*Pisum sativum*), failed to discover the laws of inheritance, with one notable exception. That was the work of the Moravian priest Gregor Mendel in the garden of the Augustinian monastery of St Thomas in Brünn (now Brno in the Czech Republic). Mendel grew 34 cultivars of the garden pea in 1854 and again in 1855 before embarking on 8 years of hybridization experiments which involved the raising of over 10,000 plants. He presented his results in a paper read twice to the Brünn Natural History Society in 1865, which they published in their proceedings in 1866. Mendel's work went unappreciated until independently 'rediscovered' in 1900 by three European botanists, de Vries in Holland, Correns in Germany, and Von Tschermak in Austria (Carlson 2004). An English translation was made by the Royal Horticultural Society of London and published in volume 26 of the Society's Journal in 1901. Mendel's paper (Mendel 1865) is well worth reading both for its historic value and for his attention to detail. He succinctly describes how to do artificial fertilization in the garden pea: the bud is opened before it is perfectly developed, the keel is removed, and each stamen carefully extracted by means of forceps, after which the stigma can at once be dusted over with the foreign pollen.

## Darwin and the Evolutionary Significance of the Mating System

Darwin studied the effects of cross- and self-fertilization in 57 species of flowering plants from 52 genera and 30 families (Darwin 1876). He did 11 years of experiments involving up to ten successive generations of self-pollination and the growing of offspring in pots in his greenhouse or in rows in his garden. The experiments were the forerunner of those done on breeding and mating systems in the twentieth century which established modern scientific breeding methods. His most important conclusion was that cross-fertilization is generally beneficial and self-fertilization

(inbreeding) injurious. This was shown by the difference in height, weight, constitutional vigour, and fertility of the offspring from crossed and self-fertilized flowers, and in the number of seeds produced by the parent-plants. Darwin did find that some species, such as the garden pea, were tolerant of inbreeding, but crosses between cultivars displayed hybrid vigour; a topic we shall discuss in detail in Part III.

The conclusion that cross-fertilization is generally beneficial provided an explanation for cross-pollination being encouraged in hermaphrodite species, despite the fact that they could reproduce with greater certainty and less pollen production by self-fertilization. The following mechanisms promote cross-pollination: two or more distinct forms of hermaphrodite flowers adapted for reciprocal fertilization (e.g. heterostyled plants), pollen shed before stigma receptive (protandrous flower), stigma receptive before anthers open (protogynous flower), and what we today call self-incompatibility (flower can be fertilized by pollen from a different plant but not by its own pollen). Furthermore, some species are monoecious or dioecious. Darwin was particularly interested in the evolutionary advantages and disadvantages of the variation found in breeding and mating systems and how one system could evolve into another, such as the shift from outcrossing to selfing (Darwin 1876, 1877). Darwin also considered the advantages and disadvantages of wind pollination (anemophilous) compared with the more efficient pollination done by pollen-devouring and nectar-seeking insects (entomophilous). Today honey bees are probably the best known pollinators and in the USA alone they have an estimated worth of around \$200 billion a year for the work they do in pollinating crops (Benson 2012). Interestingly, flies and beetles are thought to have been the original pollinators when flowering plants first emerged 140 MYA, but today butterflies, moths, bees, hoverflies and mosquitoes are all pollinators, along with hummingbirds, snails and slugs, weevils, and indeed mammals (Benson 2012).

The existence of diversity in breeding and mating systems suggests that no one system is advantageous in all circumstances. However, lack of understanding of the laws of inheritance prevented Darwin from putting forward entirely satisfactory explanations. Accounts of modern theories based on population genetics can be found in *Elements of Evolutionary Genetics* by Charlesworth and Charlesworth (2010) and research papers in *Major Evolutionary Transitions in Flowering Plant Reproduction* edited by Barrett (2008). As breeding and mating systems are under genetic control, they can be manipulated by plant breeders to achieve more efficient breeding methods and to produce hybrid cultivars.

## Purity and Seed Quality of New Cultivars

The isolation required to maintain the purity of new cultivars during multiplication depends on the mating system, and this was clearly understood by nineteenth century plant breeders. They found that wind-pollinated maize cultivars must to be planted separately so that they do not cross. Likewise, pure kinds of insect-

pollinated cabbages can't be raised in the same garden if they flower at the same time. Care is also required in *Brassica* species to stop the crossing of different vegetables (*B. oleracea*) and the crossing of root and oil forms (*B. napus* and *B. rapa*). In contrast, self-pollinated pea cultivars can be raised in the same garden and kept true to type for a long period because they are not crossed by insects and are perfectly fertile without their aid. The same is true of other plants such as sweet-peas and tomatoes.

Isolation distances became quantified during the twentieth century with the development of statutory cultivar registration and seed certification in Europe and North America, with the aim of providing farmers and growers with quality assurance. Today breeders can usually download the regulations from the website of their appropriate government agency; for example, the Food and Environment Research Agency in England and Wales: *Guide to Seed Certification Procedures 2012* ([www.fera.defra.gov.uk/plants/seeds/seedCertification](http://www.fera.defra.gov.uk/plants/seeds/seedCertification)). The isolation distances required for the production of basic and certified seed range from 5000 m and 1000 m, respectively, for monoecious hemp, through 1000 and 1000 m for beets, 1000 and 600 m for vegetable brassicas, 400 and 200 m for forage brassicas, and 200 and 200 m for maize, down to 50 and 20 m for self-pollinated cultivars of triticale. More generally, the standards which must be met by most kinds of seed include cultivar purity (trueness to type), analytical purity (a measure of gross contamination), freedom from weeds and high germination. The International Seed Testing Association (ISTA) was founded in 1924 in order to develop and publish standard procedures for seed testing. The International Seed Federation (ISF) was formed in 2002 through the merger of FIS (Fédération Internationale du Commerce des Semences 1924) and ASSINSEL (Association Internationale des Sélectionneurs pour la Protection de Obentions Végétales 1938), in recognition of the fact that plant breeders and seed producers had essentially become a single entity.

## Cellular Basis of Life

The nineteenth century also saw important developments in understanding the cellular basis of life, including ones that would impact on our understanding of heredity. These built on earlier discoveries made possible by the invention of the simple and compound microscopes at the end of the sixteenth century. Robert Hooke (1635–1703) used a compound microscope (more than one lens) and discovered what he called plant cells in slices of cork; although strictly speaking he had seen the cell walls of cork tissue. He included these observations in his *Micrographia* published in 1665. Shortly afterwards, using a simple microscope (single movable lens), Antonie van Leeuwenhoek (1632–1723) reported his discovery of single-celled organisms (micro-organisms) in water. However, it was in 1838 and 1839 that two German scientists, the botanist Mathias Schleiden and the zoologist Theodor Schwann, put forward a coherent but flawed theory of the cell as

the basic unit of all life. They observed, like others, that plants and animals were cellular in structure, and hence could be viewed as communities of cells. Before then, in 1831, Robert Brown (1833) had observed what he called the nucleus in the cells of both the vegetative and reproductive parts of many flowering plants, such as the leaf-cells of orchids and in the stigma and pollen of *Tradescantia* (Whitehouse 1973). It was a more or less spherical structure which usually occupied a central position in the cell cytoplasm. Subsequently, clear and increasingly detailed pictures of cellular organization became available in favourable organisms as advances were made in the quality of microscope lenses and in staining technology. Nuclear and cell division were described by botanists such as Wilhelm Hofmeister in 1848 (Blackmore 2012), including the fate of the double threads seen in the nucleus in a network of fibre-like material called chromatin. The threads were named chromosomes by Heinrich Waldeyer in 1888 (Carlson 2004), and are in fact the physical basis of heredity.

Today the cellular structure of plant roots, stems, leaves, flowers and seeds can be seen in great detail with more powerful optical microscopes and electron microscopes. Modern compound light microscopes still comprise the basic elements of eyepiece, objective set of lenses above specimen on platform, and illumination from below using condenser lenses; but the eyepieces are binocular and the image goes simultaneously to the operator and a digital camera. Dark field and phase contrast illumination are available as well as Köhler bright field illumination. Other advances in optical microscopes are low resolution stereo (dissecting) ones, for observing specimen (tissue) surfaces, and confocal laser scanning ones which can produce a three-dimensional image of processes occurring in living cells. The resolving power (resolution of fine detail) of the optical microscope is limited by the wavelength of light. Hence further advances had to await transmission electron microscopes, which make use of the shorter wavelengths of a beam of electrons, and became available from 1939. Cell organelles, and their structure, can be seen with the electron microscope, including chloroplasts which are the sites of photosynthesis, mitochondria which are the sites of respiration, and ribosomes which play a central role in protein synthesis. Subsequently, from 1965, the scanning electron microscope allowed more detailed examination of specimen (cell and tissue) surfaces. Excellent pictures can be found in modern textbooks of botany (Moore et al. 1998) and plant biochemistry (Heldt and Piechulla 2011) and in *Green Universe* by Blackmore (2012). *Green Universe* also contains a brief history of microscopy which includes specimen preparation: from sectioning with a razor blade, mounting on a glass slide in liquid medium and covering with a glass slip; through embedding in paraffin wax and cutting with a microtome; to the sophisticated ultramicrotomes and immunogold labelling used for electron microscopy.

## Cell Division

The type of cell division (strictly nuclear division) that occurs during plant development from a fertilized egg (zygote) is called mitosis. A copy of each chromosome in the cell is made before the onset of division so that when the process begins double threads can be seen; in fact sister chromatids joined at what is called a centromere. During division sister chromatids separate and move apart into what become two new nuclei in two new daughter cells. Thus each new cell contains the same genetic material, and this is also true for the regrowth of individuals (e.g. following cutting or grazing of pasture) and for new individuals produced vegetatively or asexually. Such reproduction may occur naturally through the production of bulbs (e.g. daffodil, onion and tulip), crown offshoots (asparagus), corms (crocus and gladiolus), rhizomes (iris and turf grasses), runners (strawberry), stolons (blackberry) and tubers (potato); all of which are modified stems. Some modified roots, such as those of sweet potato and dahlia, can also propagate vegetatively because they produce adventitious shoot buds. Vegetative reproduction can also occur through human intervention, by taking stem and root cuttings, layering or grafting.

The type of cell division (strictly nuclear division) that occurs during the formation of the male and female reproductive cells (sperm and egg) is more complicated and is called meiosis. Chromosome doubling is followed by two nuclear divisions in quick succession which result in four products, and hence a halving of chromosome number in the resulting sperms and eggs. It was Walter Sutton who in 1903 (Sutton 1903) correctly explained Mendel's laws in terms of chromosome behaviour (pairing and separation) during the two nuclear (cell) divisions of meiosis. The foundations had been laid for the development of scientific breeding methods, the subject matter of this book.

# **Chapter 2**

## **Scientific Breeding in the Twentieth Century and Future Goals**

### **Genetics and Cytology (Cytogenetics): The Foundations of Scientific Breeding**

The rediscovery in 1900 of Mendel's work of 1865 marked the birth of modern genetics. It also led to scientific breeding methods based on a sound knowledge of the inheritance of economically important traits. At the same time, cytologists were on the verge of confirming that chromosomes were the physical carriers of the genetic information. It is the genetic makeup of the products of meiosis, and the way in which they are combined by the mating system, that determines the genetic variation available to plant breeders for selection. In other words, these processes determine the new combinations of naturally occurring genetic variants (mutations) produced by sexual reproduction, and from these combinations new cultivars are selected which are superior to existing ones and which lack any serious defects. Advances in genetics and cytology (cytogenetics) during the first half of the twentieth century therefore informed the development of new plant breeding methods, including the use of genes from the wild relatives of cultivated plants. These methods have seen the replacement of variable landraces and open-pollinated strains with higher yielding and more uniform cultivars; the type determined by the mode of reproduction (sexual or asexual) and mating system (self- or cross-pollination) of the cultivated plant species.

### **UPOV**

The research work was done primarily in universities and newly founded government funded research institutes. Breeding was also undertaken by these organizations as well as commercial companies, the latter benefiting from the research. However, the real stimulus for commercial breeding came in 1961 with the adoption

in Paris of the International Convention for the Protection of New Varieties of Plants (UPOV, from the French name) which led to Plant Breeders' Rights and Royalty Income, usually lasting 30 years, in most industrialized countries. In the United Kingdom, for example, Plant Breeders' Rights were established through the 1964 Seeds Act. Unauthorized commercial propagation of cultivars was forbidden, with two exceptions. Farmers retained their right to sow seed saved from their own crops and breeders could use protected cultivars as parents for further breeding. However, the 1991 UPOV Act introduced some significant changes. It restricted the right of farmers to freely use seeds or other propagation material for further cultivation by making it an optional exception that is limited to countries which make it a special provision. Plants, and in some countries cultivars, could now be patented in addition to having Plant Breeders' Rights. If a farmer sows his or her field with a protected variety (cultivar) without paying the royalty fee, the breeder can now claim ownership of the harvest and the products made from the harvest. Finally, and ambiguously, anyone using a protected variety in breeding has to make major changes or else the 'new' variety will be considered an 'essentially derived' variety, falling to the ownership of the first breeder. The intention is to discourage small and insignificant changes in the variety's characteristic from being passed off as true innovation. UPOV headquarters are in Geneva (Switzerland) and by December 2011 there were 70 member countries ([www.upov.int](http://www.upov.int)). The use of protected cultivars as parents for further breeding is considered very important for future progress, and hence the use of patents to protect biotech crops (see later) and prevent this from happening is a cause for concern.

## Continuous Variation and QTLs

One apparent problem in the development of scientific breeding methods was quickly solved. Most economically important traits displayed continuous variation and hence individual genes affecting quantitative traits could not be recognized by Mendelian analysis because discrete classes could not be identified. Nevertheless, the foundations of our modern theory of quantitative inheritance were quickly laid by the experimental work of Johannsen (1909), Nilsson-Ehle (1909) and East (1915), and by the theoretical work of Fisher (1918). Johannsen used the offspring of individuals (progeny-testing) to show that heritable and non-heritable factors were jointly responsible for quantitative variation in seed weight of beans (*Phaseolus vulgaris*). Nilsson-Ehle discovered three Mendelian factors for red versus white grain in wheat (*Triticum aestivum*) and showed that the degree of redness was associated with the number of factors present. He and East independently realized that the segregation of a number of Mendelian factors (genes) of similar and cumulative action, together with the effects of non-heritable factors, could account for continuous variation. East used this new multifactorial (polygenic) hypothesis to explain the variation seen in the families and generations derived from a cross between two varieties of tobacco (*Nicotiana longiflora*) that

differed in corolla length. Fisher showed that genetic information could be inferred from measurements on related individuals using biometrical models which assumed Mendelian inheritance. Subsequently this biometrical approach was refined to take account of all of the properties of genes known from Mendelian genetics. It has provided the foundation for efficient conventional breeding which is still the main route to new cultivars. The concepts of heritability, additive and non-additive genetic variation, genotype  $\times$  environment interaction and population improvement are all important in predicting and improving the response to selection and hence rate of progress; as will be seen in later chapters. Today genes affecting quantitative traits can be detected indirectly through associations between trait scores and molecular markers with known positions on chromosomes, and are referred to as Quantitative Trait Loci (QTLs).

## Polyploids

Classical cytogenetics (1900–1953) revealed that many cultivated plants are polyploids, having more than the two sets of chromosomes found in the fertilized eggs of diploid species, where one (haploid) set comes from the pollen and the other homologous set from the unfertilized egg. Meyer et al. (2012) examined 203 global food crops and found that 64 % were diploids, 17 % were polyploid and 19 % had both diploid and polyploid cultivars. This last proportion was slightly larger but comparable to the number of angiosperm species that include multiple ploidy levels. Furthermore, Meyer et al. (2012) found that 37 angiosperm crops (19 %) underwent polyploidization during the process of domestication; mainly, but not exclusively perennial crops. This frequency was only slightly higher than the 15 % observed in speciation events among angiosperms, suggesting that ploidy changes do not distinguish evolution under domestication. They also found that 51 % were the result of autopolyploidy (chromosome sets from one species), 24 % were the result of allopolyploidy (chromosome sets from more than one species), while 10 % may have arisen from a combination of both. The remainder were of unknown origin or the result of ploidy reduction. Vegetatively propagated crops domesticated either for edible vegetative tissue or for fruit, but not seed crops, are most strongly associated with ploidy changes.

Polyploidization involves chromosome doubling. Although this occurs naturally at a low frequency, plant breeders developed methods to increase its frequency, particularly the use of colchicine from 1937 (Blakeslee and Avery 1937). As a result, well established polyploid species were re-synthesized to widen their genetic base for plant breeding, and new auto- and allopolyploids were created. Two well known examples of the latter are the new crops triticale (Mergoum et al. 2009b) and raphanobrassica (McNaughton 1973), both of which are crosses between two cultivated species. Triticale is the hybrid between wheat and rye and raphanobrassica is the hybrid between kale and radish.

## Mutagenesis

The mutagenic effect of X-rays on genes was confirmed in the fruit fly *Drosophila melanogaster* by Muller in 1927 (Muller 1927) and in higher plants (maize and barley) by Stadler the following year (Stadler 1928a, b). It was Stadler's research that laid the foundations for successful mutation breeding in plants. He showed that dry seeds can withstand 15–20 times higher doses of radiation than germinating seeds but that irradiation of germinating seeds gave eight times higher mutation frequencies. Furthermore, he found that the mutation rate was proportional to the treatment dose (van Harten 1998). Subsequently gamma-rays became the most widely used physical mutagen.

Attempts to obtain artificially induced mutations by treating plant seeds with different chemicals started early in the twentieth century, but the first reports of real success came in the 1940s (van Harten 1998). For example, Gustafsson and MacKey (1948) proved that mustard gas was mutagenic in barley. However, being a bifunctional alkylating agent (has two reactive groups and hence can link bases) it was particularly disruptive in its effects. Hundreds of chemicals were subsequently tested, but mutation rates remained low because the toxicity of the mutagens prevented treatment with high concentrations. The most significant development was around 1960 when a number of researchers tested the mutagenic effect of ethyl methane sulphonate (EMS) on crop plants. This monofunctional alkylating agent (has a single ethyl group to donate) went on to become the most frequently and universally used chemical mutagen in plant breeding (van Harten 1998).

Since 1960 assessments of the possibilities and limitations of mutation breeding have taken place in many crop species. These were helped by the establishment in 1964 of internationally co-ordinated research programmes by the joint FAO/IAEA Division of Nuclear Techniques in Food and Agriculture; IAEA being the International Atomic Energy Agency based in Vienna ([www-naweb.iaea.org/na](http://www-naweb.iaea.org/na)). In 2012, the FAO/IAEA database listed 3200 officially released cultivars from mutation breeding in over 200 plant species. Today mutation breeding can be an attractive option for the induction of variation for a specific trait in a given crop, particularly in successful cultivars of that crop. The aim is to produce a new cultivar more quickly and efficiently than could be done through sexual hybridization, particularly when the desired trait would need to be introduced from wild relatives of the crop species. The outcome may be a new cultivar in its own right or a parent for use in hybridization programmes. Advances in molecular biology coupled with the sequencing of plant genomes are now opening up new possibilities for mutation breeding, including TILLING (Targeting Induced Local Lesions IN Genomes) and targeted genetic alteration, as we shall see later in this book.

## DNA and Plant Genomes

The period up to 1953 is now often called Classical Genetics, 1953 being the year in which James Watson and Francis Crick published the molecular structure of DNA (Deoxyribose Nucleic Acid or Deoxyribonucleic Acid), the chemical carrier of genetic information (Watson and Crick 1953a, b), having first announced their discovery in the Eagle Pub, Cambridge (Fig. 2.1). DNA has a double helix structure in which the backbone comprises repeats of deoxyribose (a sugar) linked to a phosphate group. One of four bases, adenine (A), thymine (T), guanine (G) or cytosine (C) is attached to each sugar and faces inwards in such a way that adenine is opposite (hydrogen-bonded to) thymine and guanine is opposite cytosine (base pairs). Thus each DNA strand can act as a template for the synthesis of a complementary strand (replication of DNA) during cell division, and the order of bases in each strand is how the genetic information is coded, as we shall see later.

The prerequisite for sequencing DNA was the ability to produce a large quantity of the particular stretch of DNA of interest, in other words to clone the stretch of DNA. Herb Boyer and Stanley Cohen had developed the relevant technology for this and for genetic engineering by 1973 (Watson 2003). The desired piece of DNA is inserted into a bacterial plasmid (an autonomous loop of DNA), which in turn is inserted into a bacterium for multiplication, using restriction enzymes to cut the DNA at specific sequences of bases and ligases to join the ends of cut DNA. The first methods for reading the sequence of DNA were then developed during the second half of the 1970s by Wally Gilbert (Maxam and Gilbert 1977) and Fred Sanger (Sanger et al. 1977). Details of all these techniques and references to the original scientific literature can be found in books such as *Principles of Gene Manipulation and Genomics, Seventh Edition*, by Primrose and Twyman (2006).

The first plant genome to be sequenced was that of *Arabidopsis thaliana*, a model plant for research. The sequence published in *Nature* on 14 December 2000 covered 115.4 Mb (megabase pairs) of the 125-Mb genome and comprised an estimated 25,498 protein-coding genes. The first crop to be fully sequenced was rice, with the sequence published in *Nature* on 11 August 2005 covering 95 % of the 389-Mb genome and comprising an estimated 37,544 genes. A steady stream of sequences of other important cultivated plants has followed, including the 475-Mb genome of grapevine in *Nature* on 26 August 2007, with a prediction of 30,434 genes; the 2.3-Gb (gigabase pairs, i.e. 2.3 billion) genome of maize in *Science* on 20 November 2009, with a prediction of 32,540 genes; 86 % of the 844 Mb-genome of the potato in *Nature* on 14 July 2011, with a prediction of 39,031 genes; 760 Mb of the 900 Mb-genome of the tomato in *Nature* on 30 May 2012, with a prediction of 34,727 genes; 225 Mb of the 265 Mb-genome of the peach in *Nature Genetics* on 24 March 2013, with a prediction of 27,852 genes; 567 Mb of the 731 Mb-genome of sugar beet in *Nature* on 23 January 2014, with a prediction of 27,421 genes and 79 % of 3.48 Gb-genome of the cultivated pepper *Capsicum annuum* and its wild progenitor (*C. annuum* var. *glabriusculum*) in *PNAS* on 3 March 2014, with a prediction of 34,476 genes. Even the massive 17 Gb genome of bread wheat is



**Fig. 2.1** Molecular structure of DNA (deoxyribose nucleic acid), the carrier of genetic information, was announced by James Watson and Francis Crick in the Eagle Pub, Cambridge (England) on 28 February 1953 and 60 years later inspired a floral display at Floriade, Commonwealth Park, Canberra (Australia)

now tractable, chromosome by chromosome, as a result of technological advancements in flow-sorting of chromosomes (Eversole 2012). A draft sequence was published in *Science* on 18 July 2014 and the complete sequence is anticipated within three years. It is already clear that the number of protein-coding genes is not substantially higher in plants with large genomes, something we shall discuss later in this book. The sequencing of plant genomes should aid gene discovery and hence crop improvement. Knowledge of complete and partial genome sequences already aids the development of molecular markers and techniques such as TILLING for use in breeding, and new biotechnologies are in the pipeline (Lusser et al. 2012).

## Genetic Transformation

The years since 1953 have seen the development of DNA based molecular plant breeding methods, including genetic modification based on genetic transformation. It was studies on the chemical nature of the substance inducing transformation of types of pneumococcal bacteria that provided the first proof that DNA carried genetic information (Avery et al. 1944). Furthermore, it was research on non chromosomal antibiotic resistance that led to the discovery of genetic transformation of the bacterium *Escherichia coli* by a plasmid and the birth of genetic engineering (Cohen et al. 1972). Since the early 1980s the introduction and stable incorporation of foreign DNA into crop plants has been achieved through a number of methods, such as bombardment with DNA-coated particles, electroporation and use of the natural gene transfer system of the soil bacterium *Agrobacterium tumefaciens*.

The years since 1953 have also seen tremendous progress in understanding how genetic information is expressed (transcribed and translated) during cell division, growth and differentiation; and hence how plants utilize energy and materials from their environment for growth, development and reproduction.

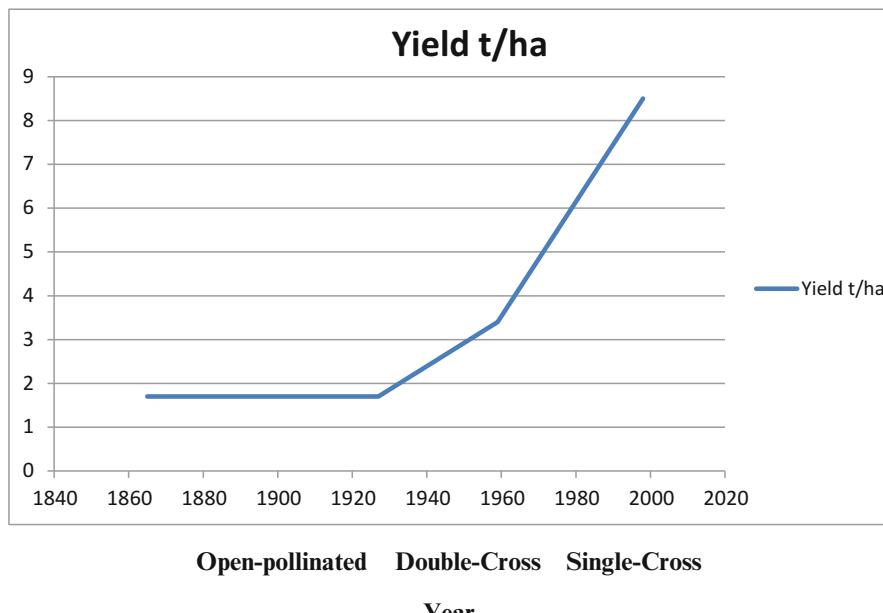
## Achievements in the Twentieth Century

Much progress was made during the twentieth century in producing new cultivars of many crops with higher yields, improved quality and better resistance to abiotic and biotic stresses. The recently published volumes (1–9 so far) in the *Handbook of Plant Breeding* (Prohens et al. 2008) record an impressive set of achievements in a wide range of crops. Here, however, I will focus on just four examples: the commercial success of hybrid maize in the USA; the Green Revolution and the prevention of famine in India; increased potato production in China and India; and the advent of genetically modified crops. The roles of international research centres and participatory plant breeding are also introduced.

## Hybrid Maize

Hybrid maize is one of the great success stories of plant breeding in the twentieth century and was the result of a combination of basic and applied research. Yields had been virtually static in the USA from 1865 to 1927, with the highest average around 1.88 t/ha (Troyer 2004). Then dramatic and steady increases occurred which saw yields reach 8.78 t/ha by 1995 (Fig. 2.2). Open-pollinated cultivars were replaced by double-cross hybrids and these in turn were replaced by single-cross hybrids. Improvements were also made in maize agronomy and these contributed about half of the increase in yields (Troyer 2006). Key changes in maize growing from 1950 to 1990 were as follows (Troyer 2004). Plant densities doubled to 74,000 plants per hectare, with row width narrowing in the 1960s from 102 to 76 cm. The herbicides 2, 4-D and Atrazine® were first commercially used in 1954 and 1965, respectively. Farmers were using about 50 kg/ha of nitrogen fertilizer in 1960, but as prices fell during the period to 1973 they applied more until applications levelled off at about 146 kg/ha in 1980. Other trends were earlier planting, the use of better pesticides and the use of larger equipment for more timely applications. The period from 1995 to 2005 saw the advent of genetically modified maize (see below) and further increases in yield to around 10 t/ha (Troyer 2006).

The science behind hybrid maize will be explained in detail in Chap. 12; here we will consider just the key stages of the story. East (1908) and Shull (1908, 1909) showed that homozygous and homogeneous pure lines (inbred lines) could be



**Fig. 2.2** Maize yields in the USA from 1865 to 1995 in t/ha (Troyer 2006)

developed from an open-pollinated maize cultivar by 5–7 generations of self-pollination. The inbred lines were often weak and difficult to propagate, a phenomenon known as inbreeding depression. When certain of the weak inbred lines were crossed, vigour was restored (hybrid vigour) and the yield of the inbred line crosses (single crosses or F<sub>1</sub> hybrids) usually exceeded that of the original open-pollinated cultivar from which the lines were developed, a phenomenon known as heterosis. The difficulty of propagating weak inbred lines appeared to prevent commercial exploitation. However, Jones (1918) solved the problem by crossing two vigorous single crosses to produce double-cross (DC) hybrid seed for farmers to grow, and showed that these too were superior to their open-pollinated cultivars. His first hybrid, made in 1917 and grown commercially in 1921, was a Burr-Leaming double-cross with two inbred lines derived from Chester Leaming and two from Burr White. These were the two highest yielding open-pollinated cultivars being grown in the 1890s. Female plants were produced by detasseling in the seed production fields in which four rows of the female parent were grown to one row of the male parent, which had to be physically removed after pollination.

Hybrids were an attractive commercial proposition because they don't breed true and hence farm saved seed is not an option. Thus from around 1922 a whole industry grew up to develop inbred lines from the best open-pollinated cultivars, and to find the best combinations for hybrid production. Funk farms released their first double-cross hybrid cultivar in 1925 and the Hi-Bred Corn Company (later Pioneer) was founded by Henry A. Wallace in 1926 (Kingsbury 2009). The period from 1933 to 1950 saw an increase in double-cross hybrids in the USA from less than 1 % of the acreage to over 99 %. After 1940, the breeding emphasis switched to crossing the elite inbred lines now available in order to make further progress. Eventually the inbred lines were strong enough for commercial production of single-cross (SC) hybrids and from 1960 these started to replace the double-cross hybrids. Also by the 1960s, detasseling had been widely eliminated by the use of cytoplasmic male sterility (CMS), discovered back in 1933. However, reliance on a single source of CMS ('Texas' cytoplasm from the open-pollinated cultivar Mexican June) proved unwise when corn leaf blight disease struck in 1970; a lesson that was learnt. So too were the wider implications of a narrow-genetic base in plant breeding, and The International Board for Plant Genetic Resources (IBPGR, now Biodiversity International) was established in 1974. By 1990 single-cross hybrids accounted for over 85 % of production in the USA, thus making Shull's original ideas a commercial reality.

## The Green Revolution

The term 'Green Revolution' was first used in March 1968 by William Gaud, the director of the United States Agency for International Development, and is associated with wheat breeder Norman Borlaug who won the Nobel Peace Prize in 1970 (Kingsbury 2009).

## ***Wheat in Mexico and India***

Borlaug had become involved with the Mexican Agricultural Program in 1943, the year it was established by the American Rockefeller Foundation and the Mexican government for humanitarian and political reasons after visits to Mexico by Henry A. Wallace, the vice-president of the USA. Borlaug took over the wheat improvement programme in 1945. Although maize and beans were Mexican staples, a rising middle class wanted wheat which had to be imported. There were three key components to Borlaug's programme. Firstly, the use of large numbers of hybridizations combined with a search for stem rust resistance. Secondly, the use of shuttle breeding between the north and the highlands to achieve two crops (generations) a year, and new wheats relatively insensitive to day length as a consequence. This made them adaptable to different regions around the world. Thirdly, from 1953, the use of short-strawed wheats, particularly Norin 10, which came to the USA from Japan after the Second World War, and were supplied by Orville Vogel of Washington State University. During the 1950s, Borlaug's new cultivars started to have a major impact on wheat production in Mexico, particularly under irrigation on large farms in the north, and the country became a net exporter of wheat in 1958. The Mexican Agricultural Program progressed through the formation of CIMMYT in 1960, with help from the Rockefeller and Ford foundations. New semi-dwarf cultivars, 'Penjamo 620' and 'Pitic 62', became available in 1962.

Borlaug visited India in 1963, and it was his subsequent collaboration with Monkombu Sambasivan Swaminathan at the All India Co-ordinated Wheat Improvement Project that had global impact. The new wheat cultivars from Mexico (e.g. Lerma Rojo 64) were shipped to India and Mexican germplasm was subsequently used in Indian breeding programmes. As a result, there were dramatic increases in wheat production, particularly in the north (the Punjab of India and Pakistan), and India was self-sufficient in all cereals by 1974, an outcome that was almost inconceivable 15 years earlier.

## ***Wheat in China***

India rather than China is usually associated with the Green Revolution in wheat production, although average yields in China have increased from less than 1 t/ha in 1949 to 5 t/ha in 2013, with total production at 126 Mt in 2014. The recent review by Qin et al. (2015) of wheat production in China therefore makes interesting reading. They collected records from regional testing data on more than 1850 Chinese wheat varieties from the 1920s to 2014. They found that since the 1920s, average grain yield has increased annually by 1.29 % for northern winter wheat, 1.5 % for southern winter wheat and 0.52 % for (western) spring wheat, although yields of the latter peaked in the 1980s. For northern and southern winter wheat, kernel number per spike and 1000-kernel weight (TKW) had increased significantly, with no change in spike number per unit area. Breeding semi-dwarf wheat in China started in the late

1950s, with plant height successfully reduced by the 1960s. Overall, since the 1920s, average plant height of northern winter wheat has decreased from around 120 cm to stabilize at 80 cm; southern winter wheat from around 120 cm to stabilize at 86 cm; and spring wheat from around 105 cm to stabilize at 89 cm.

## **Rice in the Philippines and Asia**

It is perhaps useful to set the Green Revolution in rice production in a wider context. According to Virmani and Ilyas-Ahmed (2007), global rice production increased from 145 million tons in 1948 to 599 million tons in 2000, a 411 % increase. This was achieved through an increase in area from 87 to 154 million hectares (177 %) and an increase in productivity from 1.68 to 3.89 (231 %) t/ha. Up until 1948 rice breeding had primarily been the selection and multiplication of pure lines (pure-line cultivars) from locally grown, heterogeneous landraces. Then in 1948 the Food and Agricultural Organization (FAO) of the United Nations set up the International Rice Commission (IRC) to find ways of increasing rice production. Rice breeding perspectives became global and one outcome was an ambitious *indica-japonica* hybridization programme aimed at combining the high yield and fertilizer responsiveness of *japonicas* with the grain quality and disease and pest resistance of *indicas*. Although in the main results were disappointing, pure line selection was largely replaced by recombination breeding in which inbred line cultivars were selected from the variation generated by deliberate hybridizations. The older pure-line cultivars did, however, form a rich and diverse genetic base for recombination breeding. The year 1960 (or possibly earlier) saw the release of the first short-statured *indica* cultivar Taichung (Native)-1 from the cross between spontaneous dwarf mutant Dee-Gee-Wu-Gen from Taiwan and the tall Taiwanese cultivar Tsai-Yuan-Chung. The year 1960 also saw the foundation in the Philippines of the International Rice Research Institute (IRRI) which opened in 1962 with funding provided by the Ford and Rockefeller foundations.

As with wheat, the emphasis at IRRI was on improving harvest index (grain as percentage of plant's mass) through breeding semi-dwarf rice cultivars. A cross made late in 1962 between Peta, a tall vigorous cultivar from Indonesia, and short statured Dee-Geo-Woo-Gen (DGWG) from Taiwan, proved successful. The result was the release in 1966 of photoperiod insensitive IR8, the first semi-dwarf cultivar which responded to high levels of nitrogen fertilizer without lodging. Its high yield made it a success, despite other problems, giving farmers of irrigated lowland rice a yield advantage of 1–2 t/ha over their traditional cultivars. An account of four decades of breeding for varietal improvement of irrigated lowland rice in IRRI can be found in the review by Peng and Khush (2003). Over this period the emphasis changed from increasing yield, to multiple disease and insect resistance (IR36 was most successful cultivar), to grain quality (IR64), and then back to improving yield potential through hybrid rice and 'new plant type rice'. The latter is an ideotype including low tillering capacity and large panicles that theoretically should allow the rice plant to transform more energy into grain production through

photosynthesis. Rice hybrids were first released in China in 1976 and by 1999 occupied over half the rice acreage. They have out yielded conventional inbred line cultivars by 20 % on average. We will examine them in more detail in Chap. 14. Given that a relatively small number of modern cultivars have replaced the large number of landraces once grown, an important aspect of IRRI's work has been the collection and conservation of germplasm, with over 100,000 entries now held in its genebank (Virmani and Ilyas-Ahmed 2007).

## ***Economics and Politics***

It is important to appreciate that the 'Green Revolution' didn't simply involve the production of new high yielding cultivars. The use of irrigation, fertilizers and pesticides, combined with the work of extension services and improved availability of credit to farmers, were all part of the package of modernization of agriculture. As a consequence there was both an economic and political dimension to the 'Green Revolution'. Indeed, some view the Green Revolution as part of the political strategy of the USA during the cold war (Kingsbury 2009). In contrast to the successes of the 'Green Revolution', we now know that between 1958 and 1962 possibly as many as 45 million people in China died of starvation during Mao Zedong's "Great Leap Forward" (Dikötter 2010). He thought that he could catapult his country past its western competitors by simultaneously transforming agriculture and industry through a planned economy involving collectivization, and so create a modern communist society of plenty for all; but the result was rather different. With hindsight we can see that agricultural improvement needs a sound scientific basis, and this is what subsequently happened in China. Indeed, by the year 2000, China was the biggest producer of food in the world and largely self-sufficient. It had managed to feed 20 % of the world's population with only 8 % of the world's arable land. Furthermore, it is now estimated that China holds three-quarters of the world's reserves of rice and maize, and half of all wheat stocks (McMahon 2013). A useful review of The Green Revolution, its impacts, limits and the way ahead, has been provided by Pingali (2012). Today, as well as a worldwide need to increase food production per unit of land, there is more recognition of the specific needs of African agriculture, increasing the availability of micronutrients from crops, improving the productivity of marginal environments and dealing with climate change.

## **International Research Centres**

Since 1971, international research centres such as CIMMYT, IRRI and Biodiversity International have been supported by the Consultative Group on International Agricultural Research (CGIAR), whose aim is the eradication of hunger and poverty through research, and whose funding is heavily but not exclusively

dependent on the World Bank. In 2010, CGIAR chose Agropolis International in Montpellier, France as the headquarters for its consortium of 15 International Research Centres, and took up residence in 2011. The research centres make improved germplasm available to individual countries for release as cultivars or for further breeding through National Agricultural Research Systems (NARS).

## **Increased Potato Production in China and India**

The United Nations (UN) named 2008 as the International Year of the Potato (IYP); 2004 having been the International Year of Rice, the staple food of over half the world's population. It was in recognition of the potato's role as a major food staple in contributing to the UN Millennium Development Goals of providing food security and eradicating poverty. The potato is the world's fourth most important food crop after maize, rice and wheat with 365 million tonnes fresh-weight produced in 2012 (<http://faostat.fao.org>). More importantly, over half of this production (222 million tonnes) is now in Asia, Africa and Latin America as a result of steady increases in potato production in recent years, particularly in China and India. Indeed, China (86 million tonnes) is now the number one potato producer in the world and India (45 million tonnes) is second, as a result of a major effort during the second half of the twentieth century to grow potatoes in winter in the Indo-Gangetic plains. The increases in production in China have primarily been through increases in the area of potatoes planted (fourfold since 1960), but accompanied by some increases in yield/ha (1.5-fold since 1960), whereas in India there have been equal contributions (3-fold and 2.5-fold, respectively) (Bradshaw and Bonierbale 2010). It is difficult to assess the contributions of new cultivars to the yield increases. Unlike maize there were no hybrids in which to exploit heterosis, and in contrast to wheat and rice, potatoes already had a high harvest index of around 0.80 (the proportion of the whole plant's dry-weight which is harvestable tuber). However, in India and southern China, new cultivars allowed the potato to be grown as a winter crop in short days in the lowland subtropics, and this was a major achievement.

## **Genetically Modified Crops**

The first genetically modified (GM) crops were planted in 1996 and occupied 1.7 million hectares of land in six countries. By 2014 these figures had risen to 181.5 million hectares planted by 18 million farmers in 28 countries, over 10 % of all cropland in the world (James 2014). Most of the area was accounted for by five countries: USA (73.1 million ha), Brazil (42.2), Argentina (24.3), India (11.6) and Canada (11.6), with developing countries growing around 53 % of the total. Over 90 % of the farmers (16.5 million) were small resource-poor farmers in developing

countries. A record 7.7 million small farmers in India, and another 7.1 million in China, planted over 15 million hectares of Bt-cotton (average per farmer: 1.5 ha in India and 0.5 ha in China), and 415,000 small farmers in the Philippines grew GM maize. James (2014) reported that in addition to economic gains, farmers benefited enormously from at least a 50 % reduction in the number of insecticide applications, thereby reducing their exposure to insecticides, and importantly contributing to a more sustainable environment and better quality of life. The main GM crops were soybean (50 % of total GM area, with 82 % global adoption rate), maize (30 % with 30 % adoption), cotton (14 % with 68 % adoption) and canola (5 % with 25 % adoption); other GM crops were alfalfa, eggplant, papaya, poplar, squash, sugar beet, sweet pepper and tomato. The two main traits were herbicide tolerance (HT, 57 % of total) and insect resistance (Bt, 15 %), and combinations of these (28 %); for example, 5.8 million hectares of HT/Bt soybean grown in Brazil, Argentina, Paraguay and Uruguay (James 2014).

In 2014, Klümper and Qaim (2014) did a meta- (regression) analysis of the agronomic and economic impacts of the most important GM crops to date, including herbicide-tolerant (HT) soybean, maize and cotton, as well as insect-resistant (IR) maize and cotton. They identified 147 original farm surveys or field trials done since 1995 through keyword searches in ISI Web of Knowledge, Google Scholar, EconLit and AgEcon Search. They chose to determine the percentage difference between GM and non-GM crops for five different outcome variables, namely yield, pesticide quantity, pesticide cost, total production cost and farmer profits per unit area. They found that on average, GM technology had increased crop yields by 22 % through more effective pest control and hence lower crop damage. At the same time, GM crops had reduced pesticide quantity by 37 % and pesticide cost by 39 %. The effect on the cost of production was not significant. Finally, average profit gains for GM-adopting farmers were 68 %. Yield gains and pesticide reductions were larger for insect-resistant crops than for herbicide-tolerant crops. Yield and profit gains were higher in developing countries than in developed countries. The authors concluded that their meta-analysis had revealed robust evidence of GM crop benefits for farmers in developed and developing countries.

Much of the work on GM crops has been done by multinational companies and the processes and products protected with patents, something that some people see as cause for concern (Robin 2008). Fears have also been raised by past litigations involving one of the main players, Monsanto (Robin 2008). These arose as a result of pollution from the manufacture of polychlorinated biphenyls (PCBs), used as coolants and lubricants, and from the contamination of hormonal herbicides (2,4-D and 2,4,5-T) with dioxins (e.g. tetrachloro-p-dibenzodioxin or TCDD for short). Nevertheless, GM crops now in the pipeline should address some of the worries about the impact of climate change on crop production and food security over the next 35 years, as well as producing more nutritious food crops such as ‘Golden Rice’, engineered to produce provitamin A. For example, in 2013 the first biotech drought tolerant maize (DroughtGard™) with ‘Event MON 87460’ was planted by 2000 US farmers on 50,000 ha and this increased to 275,000 ha in 2014 (James 2014). Monsanto has donated Event MON 87460 to Water Efficient Maize for

Africa (WEMA), a public-private partnership designed to deliver the first GM drought tolerant maize to selected African countries, starting in 2017 (James 2014). But the controversies remain, as is evident from the article in *Nature* by Gilbert (2013) on three pressing questions: are GM crops fuelling the rise of glyphosate-resistant ‘superweeds’?-(conclusion true); are GM *Bt*-cotton crops (insect resistant) driving farmers in India to suicide?-(conclusion false), and are the foreign transgenes in GM maize crops (not officially approved in Mexico) spreading into native crops in Mexico?-(conclusion unknown). Readers can study the evidence and come to their own conclusions. The point is that these examples show how blame shifts, myths are spread and cultural insensitivities can inflame debate.

## **Reconnecting Breeders with Farmers and End Users**

The development of scientific plant breeding saw the activity of plant breeding become divorced from the activity of farming and consequently transferred from farmers to specialist breeders, at least in Europe and North America. Yet success in plant breeding depends on producing cultivars that farmers want to grow because the produce of those cultivars meet the requirements of end users. Hence plant breeders and biotech companies realized the need to involve farmers and end users in deciding breeding objectives and assessing potential cultivars and biotech crops. Furthermore, the latter part of the twentieth century saw the development of what has been called Participatory Plant Breeding (PPB), where farmers (and other end users) are viewed as equal partners with scientists throughout the breeding and selection process. PPB has been initiated by international research institutes to improve the adoption of cultivars by poor farmers in developing countries and by those involved in organic farming in Europe and North America (Desclaux et al. 2012).

## **Future Goals of Plant Breeding: Quantity and Quality of Food Crops**

Further genetic improvements are no doubt possible and desirable in all of our cultivated plants, including ornamentals, but the most pressing needs are with food crops. Here the goals are to provide enough food of sufficient quality to meet human dietary needs, whilst being free of toxins, allergens (if possible) and unpleasant flavourings (Hahlbrock 2009). In other words, sufficient basic organic nutrients (carbohydrates, fats, proteins), all of the inorganic nutrients required by the body (mineral salts and trace elements), all essential organic substances (all vitamins, several amino acids and several fatty acids) and suitable roughage for the intestinal

tract (Hahlbrock 2009). Amino acids provide a good example of food quality. Twenty different amino acids are needed to build all of the types of proteins found in living systems. Plants can make all 20 but animals only some; humans being unable to make ten essential amino acids (lysine, leucine, isoleucine, phenylalanine, tyrosine, methionine, cystine, threonine, tryptophan and valine). Poor people often do not have access to adequate supplies of meat and hence depend on a diet of plants and plant products (King 2011). Cereals and pulse crops together provide a good balance of amino acids, the former being lysine deficient whereas the latter are lysine rich. Ideally the goals of plant breeding should be part of a global agricultural policy in which the choice of crop species makes the most appropriate use of the land available for crop production. The goals are different for people who are hungry, those who suffer under nutrition, and those who are overweight.

*Hunger:* World food production over the last 40 years kept pace with population growth, but did not exceed it (FAO, IFAD and WFP 2015). As a consequence, the proportion of hungry people has decreased from 24 to 13 %, but not the absolute numbers, so currently nearly 800 million people go to bed hungry each night (FAO, IFAD and WFP 2015). The problem is most chronic in South Asia and Sub-Saharan Africa. For them the priority is increasing the yield of staple food crops.

*Under nutrition:* A further billion people are thought to suffer from under nutrition; for example, a lack of vitamins and minerals. Hence the current interest in the biofortification of staple foods with micronutrients, a good example of which is the ProVitaMinRice Consortium launched by the Bill and Melinda Gates Foundation in 2003 as part of the Grand Challenges in Global Health Initiative ([www.goldenrice.org](http://www.goldenrice.org)). Banana is another staple food that would benefit from biofortification: it is largely deficient in iron, iodine and zinc (Pillay and Tripathi 2007). Interestingly, 2013 was made the United Nations “International Year of Quinoa” (<http://www.fao.org/quinoa-2013/en/>), a C<sub>4</sub> grain crop, in recognition of the role of the Andean people in preserving quinoa (*Chenopodium quinoa*) in its natural state as a highly nutritious food for present and future generations, through ancestral practices of living in harmony with nature. Its grain is gluten free, has high protein content (12.5–16.7 %), with a good balance of essential amino acids, and is a good source of minerals (e.g. iron, calcium and zinc) and vitamins of the B, C and E complex. As a staple food and cash crop, quinoa can play an important role in eradicating hunger, malnutrition and poverty in South America.

*Obesity:* Ironically, probably as many as a billion people are overweight with associated health problems, and this leads to different breeding objectives; for example, increasing the dietary fibre content of wheat grain and flour is being explored in the EU HEALTHGRAIN programme which started in 2005 ([www.healthgrain.org](http://www.healthgrain.org)). Furthermore, in developed countries like the USA and UK, a third of all food produced is thrown away, waste which is almost entirely avoidable (McKevitt and Ryan 2013).

## World Food Production in 2005–2007 and Future Trends

It is worth examining in more detail world food production by region at the beginning of the twenty-first century and likely trends. McMahon (2013) summarized the FAO database of food calories in the period 2005–2007 under five categories, and made predictions of likely future changes.

1. Established food powers (exporters) with a total population of around 500 million people, including the USA, Canada, Australia, New Zealand (milk from pasture) and France. Over-production of staple crops was partly ‘solved’ by feeding them to animals. Farms are large, and mechanized, heavily capitalized and high-tech. North America is expected to continue to produce a massive food surplus despite reduced rainfall in parts of the American West. There is still a large reserve of fertile land that can be brought into cultivation, and there may still be room to improve crop yields. Australia and New Zealand are also likely to grow their exports as more land is converted to agriculture, despite a drying climate.
2. Emerging food exporters with a total population of around 700 million people, including Latin American countries such as Brazil, Argentina, Uruguay, Paraguay and Bolivia; Asian countries such as Thailand, Vietnam and Myanmar; and former countries of the Soviet Union such as Russia, Ukraine and Kazakhstan. These countries have adopted and adapted many of the farming innovations of the established food powers. The Asian countries may struggle to increase production, as there is little new land available for cultivation. In contrast, the countries of the former Soviet Union have more potential as they could bring millions of hectares of fertile land into production and double their grain yields. The region with the greatest potential is South America, with its vast amounts of suitable land in reserve, even without touching the rainforests. Yields are still only half of what is physically attainable given the soils and climate.
3. The (barely) self-sufficient with a total population of 3.4 billion people, including the populous countries of Asia: China, India, Pakistan, Bangladesh and Indonesia. These countries have also adopted and adapted many of the farming innovations of the established food powers; but farms are small, labour-intensive and less heavily capitalized. China will probably need to import more food, not because it is starving as a result of serious environmental problems, but because its wealthier population has more expensive tastes. South Asia, including India, is of greater concern because the region’s population is set to increase from 1.7 billion today to 2.4 billion in 2050, 90 % of all the suitable land is already being cultivated (although crop yields could be doubled), and climate change is expected to have adverse effects. In other words, South Asian countries could become major food importers.
4. Rich countries with a total population of around 500 million people that import food. These include oil-rich-desert members of the Gulf Cooperation Council; Japan, South Korea and the UK whose populations outgrew their available farmland; and today the European Union as a whole imports (e.g. soybeans

and their derivatives) more calories than it exports as wheat, and should be regarded as a wealthy, densely populated region that relies on imported food. Japan and South Korea will remain major buyers on world markets. The position of the European Union in the global food system will probably not change much overall. It will continue to export small amounts of wheat, meat, dairy and luxury foods, while sucking in large shipments of grains, oilseeds and other raw commodities.

5. Poor and food insecure countries with a total population of around 1.4 billion people, with the largest concentration and most extreme examples found in Sub-Saharan Africa. In addition, Egypt, once the breadbasket for ancient Greece and Rome, is now a large importer of wheat as a result of a rapid population increase from 28 million in 1960 to 82 million today. The Arab World of North Africa, and Western Asia, will have to import more food. The population of this region is expected to grow from its current 450 million to 720 million by 2050, there is little additional land that can be brought into cultivation, and this is the most water-stressed part of the world with climate change predicted to make matters worse. Sub-Saharan Africa is the hardest to predict. On the one hand, the population will more than double from 878 million today to almost 2 billion in 2050, there will be many environmental challenges, and production will be hit by climate change. On the other hand, Sub-Saharan Africa has enormous potential to increase its food production. There is more than 750 million hectares of suitable land that could be brought into cultivation and African farmers could double or triple their yields and still be well below the theoretical maximum. There is much to be gained from helping small farmers in poor countries to produce more food.

There is going to be an ever greater divergence between where food is grown and where it is needed. Is the challenge of feeding the world simply equitable food distribution (trade), or are there other even more serious challenges ahead?

## Feeding the World's Population in 2050

Food security is one of the biggest challenges of the twenty-first century. It will be made more difficult by the other three great challenges of population growth, energy security and climate change (McKevitt and Ryan 2013). By 2050, humankind is going to consume energy at the rate of 28 TW (tera or trillion watts, or joules per second) per year compared with 17 TW in 2012 (McKevitt and Ryan 2013). At the same time, we need to achieve an 80 % reduction in the use of fossil fuels, currently 87 % of all fuels. We would then have a reasonable chance of keeping the rise in global temperatures due to greenhouse gas emissions down to less than 2 °C above pre-industrial levels, which were 0.75 °C less than today. This is the figure world leaders agreed at the UN climate talks in Copenhagen in 2009. It was thought that this level of temperature rise would most likely bring manageable impacts to

the earth's climate and keep the chance of runaway climate change acceptably low (Berners-Lee and Clark 2013). If we are content with a 50 % chance of exceeding 2 °C, we can emit no more than 3700 billion tonnes of carbon dioxide into the atmosphere. However, we have already emitted 2100 billion tonnes since the industrial revolution, which leaves just a further 1600 billion tonnes.

When Seppelt et al. (2014) analyzed peak-rate years for 27 of the world's major resources they found that 21 had experienced a peak-rate, with 16 centred on 2006. Conceptually, resource use proceeds through three phases: discovery and development, broad use, and increasing scarcity. The peak-rate year of resource use is the point in time with the maximum appropriation rate. Non renewable resources included fossil fuels and renewable ones included staple crops (e.g. cassava, maize, rice, soybeans, and wheat), combined with data on the consumption of animal products. Population growth and economic activity were considered the global drivers of resource use. The authors identified a sequence in the peak-rate years of resources associated with food production: 1950 for conversion to cropland, 1978 for conversion to irrigated land, 1983 for nitrogen fertilizer use (phosphate is yet to peak), then afterwards the years for food resources (1985 maize, 1988 rice, 2004 wheat and cotton, 2005 oil palm, 2006 cassava, 2007 sugarcane and 2009 soybeans). The authors inferred that the strategies to increase food production changed from land expansion to intensification of production. As peak-rate years have not yet occurred for non-renewable energy (coal, gas and oil), the authors concluded that sustained intensification of agricultural production is limited by land and not energy. The expansion, as opposed to the rate of expansion, of cropland has recently stabilized at its highest recorded levels of about  $1.8 \times 10^6$  ha each year. The authors thought that the peak-rate years had passed at a global level for the production of maize, rice, wheat and soybeans, because yield per hectare was stagnating or collapsing in 24–39 % of the world's growing areas. The authors did not have access to sufficient data to include water resources. Currently, however, farming consumes 70 % of the world's controlled flows of fresh water available to humans (Bell 2009); and increased competition for water resources can be anticipated between agricultural, industrial and domestic users. The same could be said of land and energy. Indeed, the spectacular increases in crop yields seen during the second half of the twentieth century were built on cheap and abundant energy, mostly in the form of fossil fuel (McMahon 2013). Hydrocarbons not only power the tractors on the farm, and the ships and trucks that carry food around the world, but they are also a major input for the production of fertilizers and agro-chemicals (McMahon 2013).

## ***Impact of Population Growth***

The global human population is predicted to continue rising from the 7.0 billion reached in 2011 to 9.0 billion by 2050 when 70 % of this population will be living in urban environments. After that there is great uncertainty about the likely number in

2100 (UN predictions made in 2011). As a consequence, world food production will need to increase between 70 and 100 % in just 40 years. Once again there are fears of population growth overtaking food supplies, as famously predicted by Thomas Malthus in 1798 in *An Essay on the Principle of Population*, when the population was only around 1.0 billion. Indeed, some consider that our population density has already exceeded its optimum and that overpopulation threatens the whole biosphere (Hahlbrock 2009). McMahon (2013) points out that we actually grow enough food today to feed 9.3 billion people; it is just that a lot is wasted and a lot more is diverted towards animal feed and biofuels. Our pattern of food use is a function of political choices and economic disparities. That is why McKevitt and Ryan (2013) argue that we need a global agricultural policy if nine billion people are really going to be fed, because we need to successfully distribute affordable food.

### ***Factors Affecting Food Production***

Increasing food production to feed 9.0 billion people is likely to be made even more challenging by a number of factors, the first consequences of which were seen in the sudden rise in the price of food in 2008 and subsequent riots in 28 nations, including some countries in the Middle East (McKevitt and Ryan 2013; McMahon 2013). In October 2012 the price of wheat and maize stood close to the peak price that had led to the riots in 2008. However, it is not yet clear if 2008 marked the start of a transition from an era of low and falling food prices to one of high and volatile prices. Factors affecting food production are as follows.

*Shifts in diet:* Urbanization and increasing wealth in developing countries (particularly China) is leading to a shift in diets to meat and dairy products which require the use of more farmland than an equivalent plant-based diet. This is because animals raised by intensive, large-scale livestock farming are mainly fed maize, wheat, barley, soybeans and conserved forages such as grasses, instead of grazing pastures. Feeding grain to animals is a poor way of turning sunlight into human food, although the conversion efficiencies (kilograms of grain to one kilogram of meat) vary considerably from grain-fed cattle at about ten, to pigs at about three (although six quoted below), to poultry which can be as low as 1.5 (Godfray 2014). If we don't alter our diets and continue as present, we will require twice as many cows, sheep, chickens and pigs by 2050 (McKevitt and Ryan 2013). The 20th December 2014 edition of the *Economist* carried an article *Empire of the Pig* (pp. 80–82). It reported that China now produces and consumes almost 500 million swine a year, half of all the pigs in the world. Average consumption has increased fivefold since 1979 to 39 kg per person a year with some industrial farms processing 100,000 pigs a year; less than 20 % of pigs now come from backyard farms with less than five animals. Production on this scale creates water and soil pollution from pig waste, and the waste also contributes to emissions of the greenhouse gases methane

and nitrous oxide. As each kilogram of pork requires six of feed, mainly processed soybean or maize, China needs imports on a vast scale that account for half of the world's feed crops. In 2010, for example, soybean imports were greater than 50 % of global markets and by 2022 China will need to import 19–32 million tonnes of maize. Already in Brazil 25 million hectares of former rainforest are used for soybean production to meet demand and Argentina has increased production fourfold since 1990, again mainly for export to China. It is also estimated that China has bought 5 million hectares of land in developing countries for feed crops.

*Biofuels:* First generation biofuels are using crops and cropland to produce energy rather than food so that even more food will need to be produced from the remaining land. McMahon (2013) reports most forecasts assume that biofuels will represent one-tenth of global transport fuels by 2030, and perhaps even one-fifth by 2050. As a result, the FAO expects that 12 % of the world's grain, 16 % of vegetable oil and one-third of all sugar will be used to make biofuels in just ten years time. Already in 2013, some 40 % of US corn, 50 % of Brazilian sugar cane and 60 % of European vegetable oil goes for transport fuels (Berners-Lee and Clark 2013). This trend should be reversed by second generation biofuels whose ligno-cellulosic material can be converted to simple sugars and ethanol (e.g. *Miscanthus*, elephant grass; switchgrass *Panicum virgatum*, a native grass of the American prairies; and giant cane/reed *Arundo donax*). Biogas production is another development in which a container replicates a cow's rumen with the digestate fermenting to produce methane. If suitably 'scrubbed', the methane can be pumped directly into a gas main, but more commonly is used to power an internal combustion engine linked to an electricity generator. In Britain, for example, hybrid rye and forage maize are grown as biogas crops as well as for forage; but perhaps more desirably, waste from sweet corn and other crops, and food waste, can also be used. Longer term, algae and seaweed may be used (Hahlbrock 2009). However, McKevitt and Ryan (2013) have concluded that solar energy is the only energy source with the potential to satisfy all of our needs, and that methanol rather than ethanol will probably be used to store it in liquid form. Furthermore, this isn't science fiction as plants provide us with a highly visible proof of concept from photosynthesis. But using photovoltaic cells to absorb light energy and then using this energy to produce methanol is a big challenge. In the meantime, nuclear, wind, wave and hydroelectric power can be used to reduce reliance on fossil fuels.

*Land area:* The potential for increases in land area used for agriculture were discussed in Chap. 1. However, the area may remain static or even decrease as a result of degradation and climate change, unless crops can be bred which tolerate the abiotic stresses associated with areas considered marginal for growing crops. Transforming more forest to farmland is certainly considered undesirable as rainforests are climate-stabilizing biotopes. The stresses can be due to drought, high temperatures, low soil fertility, salinity or aluminium toxicity in acid soils. Salinity is also a major threat to the sustainable irrigation required to maintain the yield and quality of many crops (Bell 2009). Rain and surface water dissolve the earth's natural minerals as they make their way to the sea, which is salty as a result.

Salt-degradation occurs in arid and semi-arid regions where rainfall is too low to maintain regular percolation of rainwater through the soil and where irrigation is practised without a natural or artificial drainage system. Water settles on the irrigated land and evaporates, leaving the salt which can accumulate to dangerous levels in the root zone of crops and reduce their productivity. In other words, the irrigation has the effect of stopping the minerals running off to the ocean. This is what brought the first civilizations to an end, such as the Sumarian one in Mesopotamia, and is still happening today (Bell 2009). Qadir et al. (2014) estimated that by 2014 about 62 million hectares (20 %) of the world's irrigated lands had been degraded by salt, up from 45 million in the early 1990s. The inflation-adjusted cost of salt-induced land degradation in 2013 was estimated at \$441 per hectare, yielding an estimate of global economic losses of \$27.3 billion per year. Well known salt-degraded land areas include: the Aral Sea Basin, Central Asia; the Indo-Gangetic Basin, India; the Indus Basin, Pakistan; the Yellow River Basin, China; the Euphrates Basin, Syria and Iraq; the Murray-Darling Basin, Australia; and the San Joaquin Valley, United States. Qadir et al. (2014) present selected case studies that highlight the potential for economic and environmental benefits of taking action to remediate salt-affected lands. Methods successfully used to facilitate drainage and reverse soil degradation include: tree planting, deep ploughing, mixing harvested plant residues into topsoil, and digging a drain or deep ditch around the salt-affected land; as well as the breeding and cultivation salt-tolerant varieties of crops.

Currently there is interest in the idea of Allan Savory of restoring the world's huge degraded arid lands using special cattle-grazing techniques designed to mimic the environmental impacts of herd animals such as bison and wildebeest. The idea is to ensure that animals move around in the manner of a wild herd being pursued by natural predators, grazing the land to keep it healthy but quickly moving on before it gets over-exploited. This regenerates the land with the side effect of turning it into a giant carbon pump that transfers carbon dioxide from the air into organic matter in the world's soils (Berners-Lee and Clark 2013).

*Bridging yield gap:* Substantial progress will be required in bridging the yield gap between what is currently achieved per unit of land and what should be possible in future, with the best farming methods and best storage and transportation of food, given the availability of suitably adapted cultivars. IIASA (International Institute for Applied Systems Analysis) have compared actual food production with the biophysical potential, as determined by local soils, terrain and climate. This allows them to estimate the 'yield gap', the amount by which output could be increased under best practice management and best production technologies. McMahon (2013) has summarized their results as follows. Currently overall the world achieves 55 % of attainable yields (gap of 45 %), with Western Europe and Eastern Asia close to 90 %, but South America at 38 % and Sub-Saharan Africa at just 24 %. In theory, African farmers could triple their output without putting any new land under the plough. This is why IIASA and FAO expect the amount of cultivated land in the world, despite the apparent abundance of suitable unused areas, will only

grow by 10 % between now and 2050. Instead, increased food production will overwhelmingly come from closing the yield gap on already cultivated land. Ray et al. (2013) used extensive agricultural statistics to estimate current yield increases per unit of land for maize, rice, wheat and soybean at 1.6, 1.0, 0.9 and 1.3 % per year, non-compounding rates, respectively. These figures are all below the 2.4 % required for a 100 % increase by 2050 so that this target may not be achieved.

## Sustainable Food Production

The intensification of food production needs to be sustainable; the higher yields must come with fewer negative consequences for the environment, including mitigation of greenhouse gas (GHG) emissions (carbon dioxide, methane and nitrogenous gases) and their effect on global warming. Even if steps are taken to stabilize greenhouse gases now, it is almost inevitable that temperatures will rise by at least 2 °C by the end of the century (McMahon 2013). Up until 2050 the overall positive and negative effects of rising CO<sub>2</sub> concentrations and rising temperatures on global food production may cancel out, as we will discuss later. However, if greenhouse gas emissions are not controlled, temperatures could rise by 5 or 6° and lead to serious problems in the second half of the twenty-first century. The Intergovernmental Panel on Climate Change Report published on 31 March 2014 (Climate Change: Impacts, Adaptation, and Vulnerability) concluded that global temperature increases of about 4 °C or more above late-twentieth-century levels, combined with increasing food demand, would pose large risks to food security globally and regionally, with risks to food security generally greater in low-latitude areas. The report also concluded, with medium confidence: for the major crops (wheat, rice, and maize) in tropical and temperate regions, climate change without adaptation is projected to negatively impact production for local temperature increases of 2 °C or more above late-twentieth-century levels, although individual locations may benefit. Projected impacts vary across crops and regions and adaptation scenarios, with about 10 % of projections for the period 2030–2049 showing yield gains of more than 10 %, and about 10 % of projections showing yield losses of more than 25 %, compared to the late twentieth century.

The maximum temperature will be determined principally by the accumulation of carbon dioxide, and to a lesser extent by nitrous oxide, because methane is short-lived with a half life of just seven years (Berners-Lee and Clark 2013). Nevertheless, cutting methane emissions could have a beneficial effect in the short term, of keeping the temperature rise down by half a degree by 2050 (Berners-Lee and Clark 2013). It is thought that food production could account for as much as one-third of man-made emissions (McMahon 2013). Thirteen per cent comes directly from farming activities, namely nitrous oxide in the production and application of nitrogen fertilizers, and methane from digestion of forage by ruminants, decomposition of manure and flooded paddy fields. The other 17 % comes from deforestation (carbon dioxide). In contrast, the International Energy Agency (IEA) estimates that

industry produces about 40 % of global energy-related carbon dioxide emissions; with key industries being iron and steel, cement, chemicals, paper and aluminium. Interestingly, grassland accounts for 34 % of the global stock of carbon in terrestrial ecosystems, and stores this carbon in the soil rather than the plant. Grassland is therefore seen as an important form of carbon ( $\text{CO}_2$ ) sequestration (Reheul et al. 2010).

There is certainly pressure for more efficient use of water and fertilizers in crop plants and environmentally compatible plant protection. There is a need for heat and drought tolerant cultivars in a number of countries, and fodder crops that reduce GHG emissions (i.e. methane) in beef and dairy cattle are also desirable. Increasing yield along with other crop improvements will be a challenging objective for breeders during a period of climate change, and realism will be important when defining objectives for a particular farming system, particularly in terms of available inputs. The editorial in *Nature Biotechnology* for March 2012 (vol. 30(3), p. 197) also argues that averting a global food crisis will require the deconstruction of several hurdles to the deployment of new strategies in plant breeding; in particular, deregulation of proved technologies and a shift of regulation to assessment of the crop traits themselves in order to provide investment incentives for innovation.

## Available Inputs and Their Utilization by Plants

Whatever the genetic makeup and potential of a new cultivar, one cannot ignore its requirements for plant growth, development and reproduction when considering its actual yield and quality in any given circumstances. Take the potato as an example. Its global average fresh-weight yield was still only 19.0 t/ha in 2012, but varied tremendously by country from 1 to 48 t/ha (<http://faostat.fao.org>). In many areas these yields were less than half the potential set by cultivar, length of growing season and temperatures. This was because of failure to plant disease-free seed tubers, inadequate application of fertilizers, inadequate supply of water, and ineffective control of weeds, pests and diseases. Addressing all of these factors would be of immediate benefit to farmers. This has happened through large-scale commercial (industrial) agriculture in Western Europe and North America, but what is the future for smallholder farming in Asia, Africa and Latin America? So let's finish this chapter by looking at inputs and their utilization by plants.

## Photosynthesis, Water, Temperature and Climate Change

Our modern understanding of photosynthesis started in the seventeenth century when Belgian physician Jan Baptista van Helmont demonstrated that the increase in mass of a willow sapling was not at the expense of the soil in which it was grown. It progressed in the eighteenth-century with discoveries by chemists such as

Antoine Lavoisier and Joseph Priestly, and the Dutch physician Jan Ingen-Housz who demonstrated the need for sunlight and green tissue. It progressed further in the nineteenth century when Swiss scientist Nicholas Theodore de Saussure showed that the weight gain of a plant came from water as well as carbon dioxide. Finally in the 1930s the work of Robert (Robin) Hill in Cambridge, England separated the light and dark reactions of photosynthesis and paved the way to our current understanding of the process (King 2011).

Plants need light, water ( $H_2O$ ) and carbon dioxide ( $CO_2$ ) for photosynthesis, to convert the sun's light energy into chemical energy that can be used to synthesize sugars made of carbon (C), hydrogen (H) and oxygen (O from  $CO_2$ ) atoms; a process whose efficiency is determined by temperature and carbon dioxide concentration as well as being light dependent. Plants in fact consist of 47.4 % hydrogen, 27.6 % carbon and 23.7 % oxygen (Moore et al. 1998). The capture of light (photon) energy by pigments (primarily chlorophyll-*a*) is a very complicated process involving two reaction centres and a transfer of electrons that results in the release of oxygen from water and the production of reducing power (NADPH: reduced nicotinamide adenine dinucleotide phosphate) and chemical energy (ATP: adenosine triphosphate from adenosine diphosphate). Then in the 'dark reaction' of photosynthesis,  $CO_2$  is assimilated via the Calvin cycle at the expense of the reductive power and ATP (Heldt and Piechulla 2011). The carbon dioxide and a limited amount of water are taken directly from the air by plants through small openings (stomata) in their leaves. Most of their water requirement comes from the soil and its transpiration through the plant is essential for the maintenance of plant structure and function. Photosynthesis is unavoidably linked with a substantial loss of water by evaporation from the leaves as they take in  $CO_2$ . A plant growing in temperate climates requires 700–1300 mol of  $H_2O$  for the fixation of 1 mol of  $CO_2$  by what is called the C<sub>3</sub> pathway, because fixation results first in the formation of the 3-carbon molecule, 3-phosphoglycerate (Heldt and Piechulla 2011). Furthermore, as temperatures increase above 25 °C, water loss starts to become a serious problem. However, the progenitors of crop plants such as maize, sugarcane and millet, which evolved in warm areas, often in dry habitats, some 35 million years ago (MYA), developed a way to considerably decrease this water loss through mechanisms involving a C<sub>4</sub> pathway. They are able to draw  $CO_2$  into their leaves through partially closed stomata which reduces water loss; incorporate, concentrate and store it in 4-carbon molecules; then release it when and where it is needed for photosynthesis. The details can be found in textbooks of plant biochemistry (Heldt and Piechulla 2011). The important point is that in warm climates C<sub>4</sub> plants, with their reduced water demand and their suppression of photorespiration, have an advantage over C<sub>3</sub> plants. Photorespiration is the process by which chemical energy from the light reactions of photosynthesis is eliminated when it cannot be used for  $CO_2$  assimilation. It does therefore have an important role in screening C<sub>3</sub> land plants (and their algal ancestors) from the damaging effects of ultraviolet (UV) and visible radiation. Furthermore, C<sub>3</sub> plants are photosynthetically more efficient than C<sub>4</sub> plants in cold temperatures, below approximately 5–7 °C (Willis and McElwain 2014). In contrast, their rate of photosynthesis reaches a maximum at a lower light

intensity than that for C<sub>4</sub> plants (King 2011). Pineapples and agave sisal, like cacti, can photosynthesize under severe water shortage in very dry and often hot habitats through crassulacean acid metabolism (CAM). They take up CO<sub>2</sub> only during the night when it is cool and humidity is comparatively high, and fix it in an acid which is degraded the following day to release the CO<sub>2</sub> (Heldt and Piechulla 2011). This method of photosynthesis is thought to have evolved about 130 MYA (Willis and McElwain 2014). Not surprisingly, plant physiologists, breeders and genetic engineers are trying to manipulate these fundamental processes in order to achieve higher and more stable yields (Leister 2012). Currently, for example, an international consortium is trying to increase the yield of rice through the introduction of C<sub>4</sub> photosynthesis (von Caemmerer et al. 2012). Less dramatically, but equally important, plant physiologists and breeders continue to debate optimal plant architecture (e.g. arrangement of leaves) and interior leaf design (arrangement of cells in the upper palisade and lower spongy layers, and distribution of chlorophyll molecules) for intercepting and absorbing as much light energy as possible.

## ***Climate Change***

Predicted climate change is likely to present both opportunities and challenges to plant breeders and agronomists. Two important changes to the earth's atmosphere are predicted to occur between now and 2050, with consequences for crop yields (Jaggard et al. 2010). The CO<sub>2</sub> concentration is predicted to increase from 370 to 550 ppm, having fluctuated between 200 and 300 ppm for the past 600,000 years, as revealed by the Vostock ice core data (McKevitt and Ryan 2013). This will increase the yield potential of C<sub>3</sub> crops such as wheat (15 %), rice (10 %), soya bean (15 %), sunflower (13 %), oilseed rape (13 %), potato (36 %), sugar beet (8 %) and dry bean (13 %) by increasing the rate of photosynthesis, but not that of C<sub>4</sub> crops such as maize, sugar cane and sorghum (Jaggard et al. 2010), or at least not to the same extent (Vanuytrecht et al. 2012). The high increase in potato yields is also seen in tomato where the air in glasshouses used for commercial production is enriched to greatly increase yield. Rosenthal et al. (2012) found an even larger increase for cassava. Under field conditions and fully open air carbon dioxide elevation (FACE) from 385 to 585 ppm, above ground biomass was 30 % greater after three and half months of growth, and storage root dry weight was 105 % higher (fresh weight 89 %). Furthermore, overall plant cyanide content was lower due to the dilution of nitrogen relative to carbon at elevated CO<sub>2</sub> levels. Interestingly, the photosynthetic temperature optimum of cassava is predicted to increase from 28 °C at current CO<sub>2</sub> concentrations to 31 °C at 585 ppm (Rosenthal and Ort 2012); an example of the synergistic effects of rising CO<sub>2</sub> concentrations and global warming. These results are encouraging as cassava has long been a starchy root crop that is vital to food security in Sub-Saharan Africa, as well as globally being the second most important root crop after potatoes. However, more generally, if the temperature climbs more than 3 °C, which is a distinct possibility, total global crop yields are predicted to fall

(Berners-Lee and Clark 2013). Temperatures will be supra-optimal for crop development. Furthermore, King (2011) has cautioned that limited nitrogen availability could prevent realization of all of the predicted yield increase from rising CO<sub>2</sub> concentrations. Despite the uncertainties, the challenge for plant breeders is to select new cultivars that will be higher yielding through adaptation to the changed, CO<sub>2</sub>-enriched atmosphere.

The second change to the earth's atmosphere is in the ozone concentration which is also likely to increase from 50 to 60 ppb as a result of intense industrialization (King 2011). The burning of fossil fuels releases NO<sub>2</sub> which is dissociated by sunlight into NO and O, the latter reacting with O<sub>2</sub> to yield O<sub>3</sub> (ozone) in the troposphere. Here ozone enters plant leaves through stomata and as a strong oxidant causes several kinds of damage. It can also damage guard cells so that stomata remain closed. The yields of all crops are predicted to decrease by between 4 % (rice) and 12 % (soya bean) as a result of the increased ozone concentration. Hence when the effects of CO<sub>2</sub> and O<sub>3</sub> increases are combined, potential potato yields could increase by 28.5 %, wheat and rice by 6 %, but maize yields fall by 5 % and sorghum and sugar cane by 7.5 % (Jaggard et al. 2010). In contrast, the production of ozone from the reaction of sunlight with oxygen in the stratosphere is desirable, as it intercepts ultraviolet light at 180–240 nm wavelengths. This is the part of the solar spectrum most damaging to living tissues.

## Water Use

Transpiration of water from leaves is the main driving force used to move water, and minerals, from the soil to all parts of the plant, which for trees can be to a great height. Evaporation of water has a cooling effect and complements convection to keep plants cool (King 2011). The water retained in the plant is used for photosynthesis, to maintain the turgor of living cells necessary for their functioning, and to prevent wilting of tissues and organs. Water is also the medium for cellular metabolism (chemical reactions) and comprises 70–95 % of a plant's weight (King 2011).

Increased CO<sub>2</sub> concentration increases its rate of diffusion into leaves through the stomata, relative to the rate at which water vapour diffuses out. This increases the water use efficiency (WUE) of C<sub>3</sub> plants because more dry matter is produced per unit of water transpired. It also reduces water consumption by all crops through partial closure of stomata. However, the gaseous changes in the atmosphere are predicted to change the earth's climate, making it warmer by an average of about 2 °C, even if effective steps are taken to reduce greenhouse gas emissions. This temperature rise will increase evapotranspiration rates and hence negate the reduction in water consumption. The increased evaporation of water from plants and other wet surfaces is likely to lead to increased (warmer air holds more water) but more variable precipitation, an aspect of our weather which is still predicted with a great deal of uncertainty. The Intergovernmental Panel on Climate Change (IPCC) report of 27 September 2013 predicted that annual mean precipitation is likely to

increase in high latitudes, the equatorial Pacific Ocean and mid-latitude wet regions, but decrease in many mid-latitude and subtropical dry regions. Extreme precipitation events over most of the mid-latitude land masses and over wet tropical regions will very likely become more intense and more frequent.

Despite some increases in precipitation, the most pressing constraint on food production is likely to be water scarcity (McMahon 2013). Irrigation is crucial to the world's food supply: only 17 % of the world's land is irrigated but this land produces around 40 % of the world's food. One of the world's most remarkable irrigation schemes can be found in China and dates back to 250 BCE. The Dujiangyan Irrigation System lies on the Minjiang River in the northwest of Dujiangyan City (Fig. 2.3). The river runs across the vast Chengdu Plain and caused localized flooding of settlements in ancient times. Li Bing, the local governor, came up with an ingenious solution: the construction of a dyke (levee) in the middle of the river to redirect a portion of its flow into an inner river from where excess water could be discharged into a channel cut through Mount Yulei to irrigate the dry Chengdu Plain. The outer river remained a natural course whose main function was to release the flood in the rainy season. The scheme was designed to allow 40 % of the water into the inner river in the flood season and 60 % in the dry season. Furthermore, 90 % of the silt was carried away by the outer river.

Since the 1970s, increasing amounts of irrigation water have been extracted from underground aquifers. Today 40 % of global agriculture relies on aquifers for growing crops (Bell 2009). Some of these are replenished every year while others



**Fig. 2.3** The Dujiangyan Irrigation System, Chengdu, China. (a) Looking down on dividing dyke (b) “Fish’s Mouth” divides the river (c) Inner and outer river

were formed thousands of years ago and are no longer replenished. One of the largest aquifers in the world is the Ogallala, which stretches under the Great Plains of the USA. It provides about 30 % of the irrigation water used on American farms. The water level is falling and some expect it to dry up completely in as little as 25 years. There are similar concerns on the North China Plain and in the Indian states of Punjab and Haryana. With climate change, shrinkage of montane snow fields and glaciers is expected and will lead to more variable river flows (Godfray 2014). The latter is of particular concern for the agricultural areas north and south of the Himalayas. Nevertheless, the FAO estimates that globally little more than half of the land suitable for irrigation is currently irrigated (McMahon 2013). There are opportunities to increase the supply of water to farmers by investing in dams, canals and small-scale water conservation features. Furthermore, only just over one-third of irrigation water is actually taken up by plants, the rest is lost to evaporation, run-off or ground seepage. This is because only 1 % of irrigation water goes through drip irrigation, the most efficient method by far. Desalination is another possibility for increasing the supply of water if sufficient cheap energy is available. McKevitt and Ryan (2013) reported that Sundrop Farms is using solar power in desert regions in South Australia and Qatar to desalinate seawater for irrigation and to regulate the temperature within glasshouses for growing tomatoes and peppers.

## ***Temperature***

Temperature is not an input but it affects many physiological processes that influence yield, and as mentioned above, rising temperatures from global warming are a key component of predicted climate change. The effect of temperature on crop yields is often confounded in practice with irradiance, water stress and pests and diseases. The subject was extensively reviewed by Evans (1996) in his book *Crop Evolution, Adaptation and Yield*. Just a few points will be made here. Temperature influences yield through its effect on crop growth, on the number of inflorescences, seeds or storage organs initiated, and on the rate and duration of their development. In cereals the fall in the duration of grain growth as the temperature rises is compensated to some extent by associated increases in the rate of grain growth. Thus in maize the optimum compromise was found to be around 25 °C. Some researchers have argued that tropical temperatures *per se* do not reduce yield, even for temperate crops such as wheat, but that reduced interception of radiation due to shorter life cycles is the limiting factor. However, the maps reproduced by Singh and Trethowan (2007) show by 2050 a marked increase of heat-stressed wheat growing areas in South Asia, particularly in India. Furthermore, Virmani and Ilyas-Ahmed (2007) reported predictions that yield of dry-season rice in the Philippines would decrease as much as 15 % for each 1°C rise in mean temperature during the growing season, and likewise (14.5 %) for summer rice in India. Interestingly, Evans (1996) thought that as the atmospheric CO<sub>2</sub> concentration rises, provided irrigation and fertilizer management improve, the adverse effects of high temperatures on yield in many crops could be ameliorated by selection for heat tolerance.

## Fertilizers and Nitrogen Fixation

Plants need to extract macro- and micro-nutrients (inorganic minerals) from the soil as ions in aqueous solution, unless grown directly in aqueous solutions of mineral nutrients (hydroponics). These include nitrogen in the form of nitrate (or less commonly, ammonia), from the degradation of biomass, unless the plants have a symbiotic association with micro-organisms that can fix atmospheric nitrogen. Such organisms are thought to have evolved early in Earth's history and certainly long before soils had become well established by 440 MYA (Willis and McElwain 2014). The average composition of plant tissue is 0.8 % nitrogen, 0.2 % potassium, 0.1 % calcium, 0.06 % magnesium, 0.05 % phosphorous, 0.02 % sulphur and then less than 0.01 % in total of the micronutrients chlorine, iron, boron, manganese, zinc, copper and molybdenum (Moore et al. 1998). Iron is sometimes included with the macro-nutrients as it is also an ash element, left behind when plant material is incinerated at very high temperatures. Nickel is also an essential micro-nutrient, whereas sodium, silicon, selenium and cobalt are classed as beneficial elements, although known to be essential to some plants (King 2011). Sodium, for example, is required by C<sub>4</sub> plants.

Scientific insights into the fundamentals of plant nutrition came during the nineteenth century from the work of scientists such as the German botanist Julius von Sachs (1832–1897) who developed hydroponics (King 2011), and the German chemist Justus von Liebig (1803–1873) who is considered the father of the modern fertilizer industry (Hahlbrock 2009). As a result of fertilization it was possible to intensify agricultural production on fields whose soils had been leached and depleted of nutrients. Thus the large-scale Haber–Bosch process of synthesizing ammonia (and hence ammonium salts) from atmospheric nitrogen was a major breakthrough in 1916. The provision of nitrogen fertilizer has led to huge increases in yields in developed countries; for example, about twofold for wheat in the UK with the optimum dose of nitrogen fertilizer, which is now about 200 kg/ha (Jaggard et al. 2010). Between 1950 and 1980, average N dressings for winter wheat increased from 50 to 180 kg/ha but have risen only slowly since then, because the cost of nitrogen fertilizer is an issue for farmers. Today there are also concerns over the sustainability of using fossil fuels for its production and the environmental impact of its use, namely widespread pollution of groundwater (drinking water). Bacteria in human intestines can convert nitrate to nitrite which binds to haemoglobin, particularly in foetal blood. Hence there is concern about the possibility of anaemia in very young children (King 2011). There is also concern about overuse of fertilizers leading to leaching of excess nitrate (and phosphate) into waterways, which can cause an explosive growth of algae and aquatic plants (eutrophication) resulting in a depletion of the dissolved oxygen required by fish and other aquatic animals (King 2011). Recently an analysis has been published (Howden et al. 2011) of the water quality of the river Thames in London from 1868 to 2008. Nitrate levels were 1–2 mg N/L from 1860 to 1880, increased from 2 in 1880 to 4 in 1940, rose to a plateau of 8 in 1970, and are now slightly less.

The authors attributed the largest increases to the conversion of 30 % of grasslands to arable production during World War II (land management) and to increasing applications of nitrogen fertilizer up until a peak in 1985 (farming practice). The authors also concluded that the benefits from the introduction of Nitrate Vulnerable Zones in the 1990s, to control the application of manures to land, cannot be properly assessed until about 2025 because of the time delay in having effects on the river. This study clearly demonstrates the complexity of balancing agricultural production and water quality.

Plant physiologists, breeders and genetic engineers are trying to manipulate the fundamental processes by which plants extract nitrogen from the soil and then make use of it for growth and development in order to achieve higher and sustainable yields. In other words, the aim is to increase the efficiency of uptake and the efficiency of assimilation of nitrogen into amino acids (and storage proteins) and other nitrogen compounds. Furthermore, there are ambitious projects aimed at transferring the ability of some plants (e.g. legumes) to enter a symbiosis with nitrogen-fixing bacteria to those unable to establish this relationship, like wheat, maize and rice ([www.gatesfoundation.org](http://www.gatesfoundation.org)). Interestingly, it was the Chinese over 2000 years ago, who first appreciated the role of soybean (a legume) in restoring soil fertility (McMahon 2013).

Phosphorus uptake provides an interesting contrast to that of nitrogen. Phosphate can be delivered to the soil by the weathering of rocks, but in many countries it is in short supply in the soil and hence a limiting plant nutrient (Gilbert 2009). Farmers therefore add phosphate-based fertilizers to increase their crop yields, but world reserves of phosphate-rock may last for only another 100 years. Two-thirds of global production comes from just three countries: China, the USA and Morocco. The search is on for unconventional reserves and for recycling technologies, but so too is research into phosphate uptake and assimilation by plants. Because of its low solubility, extraction of phosphate from the soil by plant roots requires a very efficient uptake system. When availability is low, most plants can increase their uptake of phosphate and other mineral nutrients (e.g. nitrate and potassium ions) through a symbiosis with fungi. For example, the arbuscular mycorrhiza has been detected in more than 80 % of all terrestrial plant species (Heldt and Piechulla 2011); hence improving the efficiency of this symbiosis is another possible approach to higher and sustainable yields. Finally it is worth mentioning that plants can synthesize all of their sulphur compounds by assimilating sulphate taken up from the environment. The details can be found in the book by Heldt and Piechulla (2011).

## Reducing Yield Losses from Weeds, Pests and Diseases

Crops often fail to reach their yield potential because of competition from weeds and losses due to attacks by pests and diseases. In Western Europe and North America these losses have been greatly reduced since the 1950s by the use of

modern herbicides, insecticides and fungicides. However, chemical control is expensive, not always effective both short-term (application problems) and long-term (evolution of resistant strains of pests and diseases), and raises environmental and health concerns, particularly over large-scale pesticide use. DDT (dichloro-diphenyl-trichloro-ethane) was first synthesized by a German chemist in 1874, but its insecticidal properties were not discovered until 1939. Almost immediately it was hailed as a means of eliminating insect-borne disease (e.g. malaria) and winning the farmers' war against the destroyers of their crops. However, as early as 1962, Rachel Carson in *Silent Spring* (Carson 1962) was warning of the dangers of the large scale and indiscriminate use of what she dramatically called the elixirs of death: insecticides such as the chlorinated hydrocarbons and organic phosphates, and herbicides such as 2,4-D (2,4-dichlorophenoxyacetic acid) and 2,4,5-T (2,4,5-trichlorophenoxyacetic acid). A 50–50 mixture of these two herbicides, known as Agent Orange, was used by the US Military from 1965 as a defoliant in the Vietnam War, but the 2,4,5-T was contaminated with a dioxin (Robin 2008). For this reason, use of 2,4,5-T as a herbicide was finally banned in the USA in the mid 1970s, and subsequently in the rest of the world. Rachel Carson was concerned about the safety of those applying the chemicals and unintended contact with the chemicals by humans and wildlife. She was also concerned about the chemicals polluting surface waters and underground seas as well as poisoning the inhabitants of the soil, which are essential for the growth of land plants, and hence all animal life. Finally, she was particularly concerned about the chemicals being passed between organisms along food chains. Fifty years on, there are continuing concerns, despite the introduction of many safety regulations. One example is the implication of neonicotinoid insecticides in the decline of bee populations, so essential for the pollination of many plant species. The levels of these compounds found in the nectar and pollen of crop plants is sufficient to reduce the growth rate of colonies of bumble bee (*Bombus terrestris*) and the production of new queens by 85 % (Whitehorn et al. 2012). Furthermore, non lethal exposure of honey bees (*Apis* species) to thiamethoxam (a neonicotinoid systemic pesticide) has been shown to cause high mortality due to homing failure at levels that could put a colony at risk of collapse (Henry et al. 2012).

Cultivars with inbuilt resistance to pests and diseases, as well as the ability to suppress weeds, would address all of these concerns. Hence these will remain highly desirable breeding goals for the indefinite future. Priorities may change as a result of climate change because pathogens and pests have specific temperature and moisture requirements for their population growth. Jaggard et al. (2010), for example, concluded that soil-borne pathogens are likely to be an increasing problem as warmer weather will increase their multiplication rates. Achieving high levels of durable resistance can be difficult, as we shall see later in this book, although new possibilities may come from transgenic approaches to breeding for resistance. Equally challenging for the breeder is combining such resistances with high yield potential and the qualities required for particular end uses.

## Goals, Selection Criteria and Success

In this chapter, major achievements over the last 115 years and future goals have been considered in broad terms, with an emphasis on increasing the quantity and quality of staple food crops. Current breeding goals in a wide range of crops can be found in the recently published volumes (1–10 so far) in the Handbook of Plant Breeding (Prohens et al. 2008). The goals are set in terms of producing new cultivars better adapted to the farming and production system under consideration, including levels of inputs such as fertilizer and water, and to the available growing season(s). They are also set in terms of meeting end user specifications. Lastly, consideration is given to the desirability and feasibility of inbuilt resistance to major pests and diseases, suppression of weeds, and tolerance of abiotic stresses such as heat and drought. The task of breeders is to translate these overall goals into selection criteria that can be used in breeding programmes. Whether these criteria are the trait itself, physiological and biochemical components of the trait, or genes and DNA sequences, will depend upon our level of knowledge and understanding of how genetic potential develops into the desired plant characteristic. Whatever the social, political and economic contexts of the programmes, understanding the nature of plant breeding will be essential for the success of the endeavours. The historical stages in the genetic improvement of the world's four most important food crops are summarized in Table 2.1. They have been introduced in these first two chapters and will be considered in detail in Parts III and IV on breeding methods. First, however, in Part II we are going to consider how useful genetic variation is generated and recognized.

**Table 2.1** Genetic improvement of world's four most important food crops (SC single-cross, DC double-cross, GM genetically modified)

Method	Wheat	Rice	Maize	Potatoes
Selection	Domestication	Domestication	Domestication	Domestication
Ploidy	Allohexaploid	Diploid	Diploid	Autotetraploid
Natural reproduction	Seeds from self-pollination	Seeds from self-pollination	Seeds from cross-pollination	Tubers from tubers or seed
Selection	Landraces	Landraces	Landraces	Landraces
Selection	Pure lines	Pure lines	Open-pollinated	Clones
Hybridization and selection	Inbreds from recombination	Inbreds from recombination		Clones from recombination
Inbreeding and crossbreeding		SC hybrids	DC hybrids	
			SC hybrids	
Introgression	New trait	New trait	New trait	New trait
Mutagenesis	New trait	New trait	New trait	New trait
Further breeding cycles	Inbreds from recombination	SC hybrids	SC hybrids	Clones from recombination
Transformation	GM cultivar	GM cultivar	GM cultivar	GM cultivar

**Part II**

**Reproduction, Heredity and Variation**

# **Chapter 3**

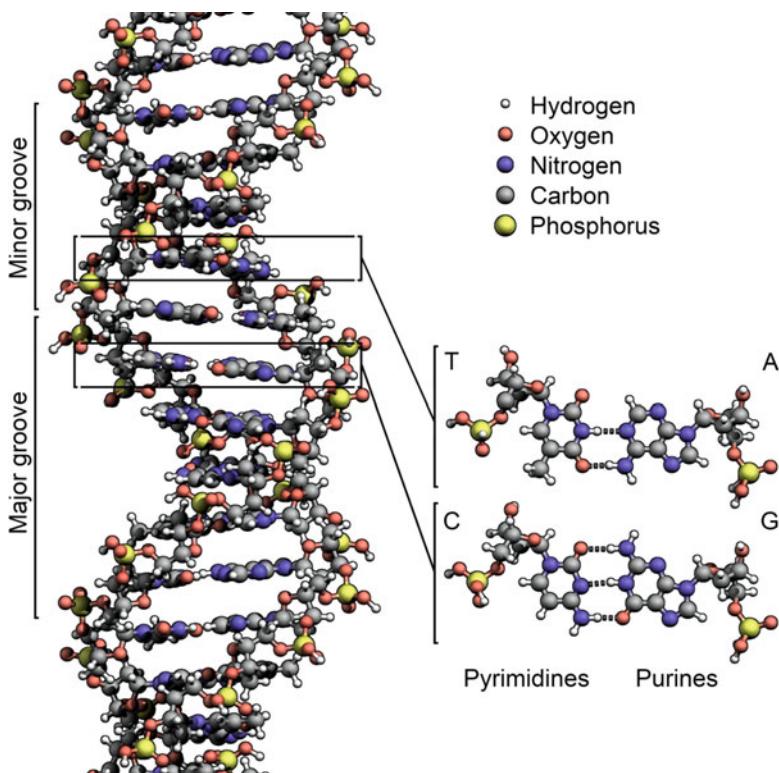
## **DNA and the Origin of Variation**

### **Introduction**

Plant breeding is the genetic improvement of cultivated plants through the production of new cultivars. As explained in Part I, the prerequisites for planned hybridizations and scientific breeding methods were an understanding of sexual reproduction and of heredity in flowering plants. Here in Part II we need to consider how useful genetic variation is generated and recognized, so that new and desirable variants can be selected and propagated. Then in Part III we will examine breeding methods and cultivar production in detail before considering the way ahead in Part IV. So let's start with DNA as the chemical carrier of genetic information and examine the life cycle of flowering plants, the importance of meiosis in combining genetic variants, and the origin of those variants.

### **The Chemical and Physical Basis of Heredity and Variation**

In the previous chapter we saw that chromosomes are the physical carriers of genetic information and their DNA the chemical basis of heredity. This remarkable macromolecule can be replicated prior to cell division (mitotic and meiotic), recombined with other DNA during meiosis and repaired after damage with various outcomes including mutation. It was studies on the chemical nature of the substance inducing transformation of types of pneumococcal bacteria that provided the first proof that DNA carried genetic information (Avery et al. 1944). Continued proof that DNA is the chemical carrier of genetic information has come since the early 1980s with every introduction and stable incorporation of foreign DNA into crop plants (genetic transformation), and since 1996 with every genetically modified (GM) crop that has been grown. DNA normally occurs in double-stranded form, the two component strands being interwound in a double helix. Each strand comprises



**Fig. 3.1** The structure of the DNA double helix: T (thymine), A (adenine), C (cytosine) and G (guanine) (Source: Wikipedia) DNA Structure+Key+Labelled.png NoBB" by Zephyris - Own work. Licensed under CC BY-SA 3.0 via Commons. [https://commons.wikimedia.org/wiki/File:DNA\\_Structure%2BKey%2BLlabelled.png#media/File:DNA\\_Structure%2BKey%2BLlabelled.png>NoBB.png](https://commons.wikimedia.org/wiki/File:DNA_Structure%2BKey%2BLlabelled.png#media/File:DNA_Structure%2BKey%2BLlabelled.png>NoBB.png)

repeats of deoxyribose (a sugar) linked to a phosphate group. The deoxyribose is able to form ester bonds with phosphate through hydroxyl groups on both carbon atoms 3 and 5; hence the 5' to 3' strand runs: phosphate-5'-deoxyribose-3'-phosphate-5'-deoxyribose-3'..., until eventually terminated by a 3'-hydroxyl group. One of four bases, adenine (A), thymine (T), guanine (G) or cytosine (C) is attached to each sugar and faces inwards in such a way that adenine is opposite (hydrogen-bonded to) thymine and guanine is opposite cytosine (base pairs) (Fig. 3.1). Thus each DNA strand can act as a template for the synthesis of a complementary strand (replication of DNA) during cell division, and the order of bases in each strand is how the genetic information is coded, as we shall see later.

The first crop to have its DNA fully sequenced was rice, with the sequence published in *Nature* on 11 August 2005 covering 95 % of the 389-Mb (megabase) genome and comprising an estimated 37,544 genes. A steady stream of sequences of other important cultivated plants has followed and these can be regarded as the physical maps of the DNA in the chromosomes of the crop species. The distances are measured in base pairs (bp or simply bases b) for the haploid genome with

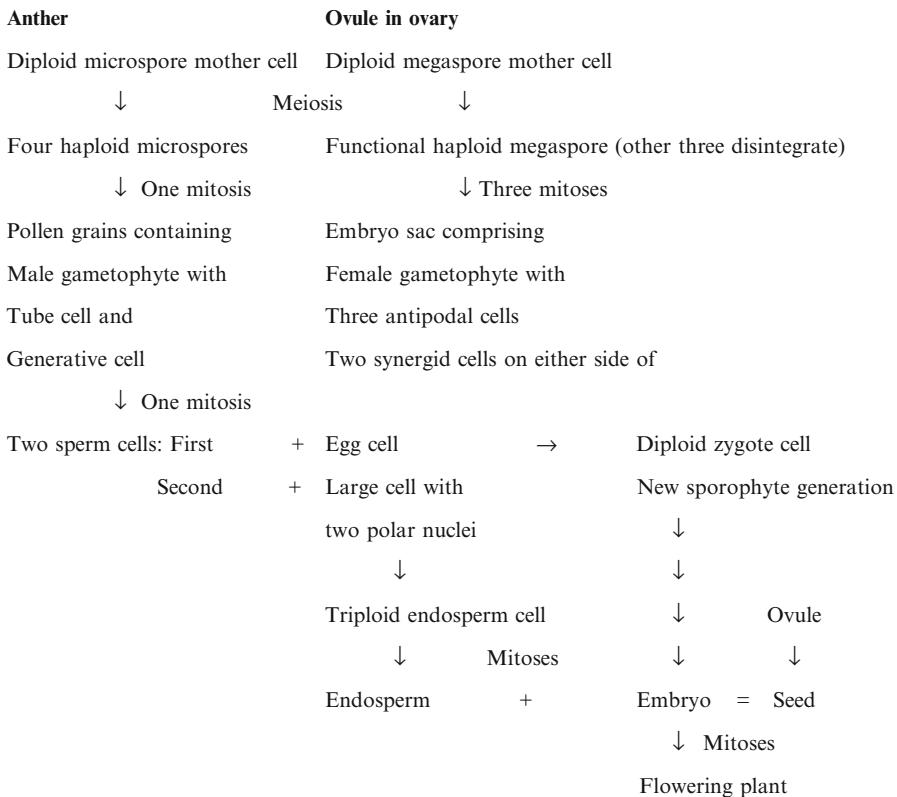
prefixes of kilo ( $\text{kb} = 10^3$  = one thousand bases), mega ( $\text{Mb} = 10^6$  = one million) or giga ( $\text{Gb} = 10^9$  = one billion). The first methods for sequencing DNA were developed during the second half of the 1970s by Wally Gilbert (Maxam and Gilbert 1977) and Fred Sanger (Sanger et al. 1977).

### ***Next-Generation Sequencing of DNA***

Since 2004 the speed at which genomes can be sequenced has dramatically increased, some would say revolutionized, thanks to what are called “Next-Generation Sequencing (NGS)” methods (e.g. Roche/454 FLX Pyrosequencer, Illumina Genome Analyzer and Applied Biosystems SOLiD™ Sequencer). Each method uses a complex interplay of enzymology, chemistry, high-resolution optics, hardware and software engineering. The bases of a small fragment of DNA are sequentially identified from signals emitted as each fragment is resynthesized from a DNA template strand, but in a parallel fashion across millions of reactions, rather than being limited to a single or a few DNA fragments. The addition of adapter sequences means that the fragments can be selectively amplified by the polymerase chain reaction (PCR) without a bacterial cloning step. The interested reader can find the details in review articles such as the one by Mardis (2008). The whole genome of an individual, or shorter sequences from a large number of individuals, can therefore be sequenced in a relatively short period of time. Thus differences between individuals of a species in base sequence can be identified. Differences at a single location are called a single nucleotide polymorphism (SNP) and these and larger differences in base sequence will be encountered throughout this book. We are now going to examine what happens to chromosomes and their DNA during the life cycles of multicellular plants.

## **Life Cycle of Multicellular Plants**

The life cycle of multicellular plants involves an alternation of gametophytic and sporophytic phases, a theory first put forward by Wilhelm Hofmeister in 1851 (Blackmore 2012). Male and female gametophytes produce sperm and egg cells (gametes), respectively, which unite to form the sporophyte that bears spore-producing cells. Gametophytic cells usually contain half the number of chromosomes found in sporophytic cells, the chromosomes being the physical carriers of heredity. For now we will consider the situation where the gametophyte is haploid, containing one complete set ( $n=x$ ) of chromosomes, the haploid genome of the plant; and the sporophyte is diploid, containing two complete sets ( $2n=2x$ ). There is tremendous variation in  $x$  between species of plants, from 2 in the flowering plant *Haplopappus gracilis* (Jackson 1957) to approximately 510 in the fern *Ophioglossum petiolatum* (Manton and Sledge 1955). The life cycle also involves



**Fig. 3.2** Life cycle of a typical flowering plant

two types of cell division, mitosis and meiosis, which are dealt with after describing the life cycle of a typical flowering plant (Fig. 3.2). The life cycles of more primitive land plants can be found in textbooks of botany; but a few key features are as follows.

The gametophyte dominates the life cycle of bryophytes, the sporophyte being short-lived and attached to the gametophyte. The bryophyte sporophyte is a simple structure with a short stalk and a rounded sporangium within which large numbers of haploid spores are produced by meiosis. These disperse and germinate to establish new green plants which are gametophytes. These have specialized structures that produce male and female gametes: the male gametes swim through a film of water to fertilize the female gametes (egg cells). Each egg cell is contained within an archegonium. Here the fertilized egg develops via an embryo into the sporophyte and the life cycle is complete.

The evolution of seedless vascular plants sees the sporophyte dominate the life cycle, but the smaller gametophyte is nutritionally independent of the parent sporophyte. Some groups of species display heterospory: the sporophyte has some sporangia that produce thousands of microspores by meiosis and others that

produce a much smaller number of megasporangia, also by meiosis. Microspores germinate to produce small male gametophytes which release the motile male gametes, whereas megasporangia germinate to give female gametophytes that develop into small plants with archegonia. Fertilization of an egg cell takes place in an archegonium and the resulting embryo grows into a new diploid sporophyte, thus completing the life cycle.

The origin of seeds can be seen in the fossil record of the now extinct seed ferns. The two living groups of seed plants are the gymnosperms with their naked seeds and the angiosperms with their seeds fully enclosed in carpels. The innovation in reproduction was the retention of megasporangia by the sporophyte instead of releasing them into the environment around the parent plant. A new plant organ (the ovule) evolved that comprised diploid tissues of the sporangium together with the fertilized female gametophyte that develops into an embryonic plant. Four groups of gymnosperms can be recognized: cycads, *Ginkgo*, the Gnetales (e.g. Ephedra) and the conifers. Cycads and *Ginkgo* have separate male and female plants and like the seedless plants require water for their motile male gametes (produced in male cones) to swim through to reach the egg. The Gnetales are a very variable group of plants and need not concern us here. Most conifers have male and female reproductive organs (cones) on the same individual plant. The male cones shed vast numbers of pollen grains which are dispersed by wind. The shape of the female cone affects the adjacent air flow and increases the efficiency with which pollen is captured from the air. In the pine family the pollen is caught in pollination droplets and only germinates to produce a pollen tube once it comes into contact with the tissues of the ovule. More generally, the key feature of conifers (and all gymnosperms) is that they produce pollen that goes directly to the ovules instead of to a stigma as in angiosperms; the stigma being the receptive region of the carpel. In conifers, after pollination the cones close to protect the developing seeds until ready for dispersal, usually by wind as winged seeds. It can be argued that the production of pollen and seeds completed the conquest of dry land, and air, by plants.

## Life Cycle of Angiosperms

Angiosperms can be classified as polycarpic or monocarpic (King 2011). The former are capable of reproducing many times during their lives and are typified by trees, shrubs, perennial grasses, and plants with bulbs or other storage organs. In contrast, monocarpics flower and set seed once in their lifetime; typically they are annuals and biennials but include a few perennials. Agricultural crops are clearly monocarpic: whole fields of plants germinate, mature, flower, set seed and die at much the same time.

## ***Pollen Grain***

An anther is made up of four microsporangia called pollen sacs. Within the developing anther two distinct lineages of cells divide by mitoses to give the inner fertile pollen mother cells and the surrounding layers of sterile cells, the innermost of which is the tapetum. A diploid microspore mother cell divides meiotically to produce four haploid microspores. As they mature, the microspores develop a highly ornamented outer protective wall. Each microspore becomes a pollen grain by a mitotic cell division inside the spore wall to give an immature male gametophyte comprising a tube cell and a generative cell. After a pollen grain reaches a receptive stigma, it germinates, and the tube cell forms a pollen tube which grows down the style to an ovule in the ovary. En route the generative cell divides by mitosis to generate two sperm cells which move through the tube to a small opening in the ovule, called the micropyle. These events were first seen in orchids, in 1884, by the botanist Eduard Strasburger (Whitehouse 1973).

## ***Embryo Sac***

The ovule comprises the gametophytic embryo sac together with some tissues of the parent plant, the diploid sporophyte. In the ovule, the diploid megasporangium undergoes meiosis to produce four haploid megaspores, three of which commonly disintegrate, but not in all flowering plants as explained in botany textbooks. The nucleus of the remaining functional megasporangium then undergoes three mitotic divisions to give eight free nuclei in a common cytoplasm. The eight nuclei are partitioned into the seven ‘cells’ of the embryo sac which is the female gametophyte. There are two synergid cells on either side of the egg cell, three antipodal cells at the opposite end of the embryo sac, and in the middle a large cell containing two polar nuclei which have migrated from each pole of the developing embryo sac.

## ***Fertilization and Development***

The two sperm cells enter the embryo sac through one of the synergids. One sperm fertilizes the egg to produce a diploid zygote cell, and the other one moves to the polar nuclei and fuses with them to produce a triploid ( $3n$ ) endosperm cell. Variants of this theme in higher plants can result in a different ploidy level in the endosperm. The diploid zygote cell is the first cell of the new sporophytic generation and develops by mitotic cell divisions into the embryo. The endosperm cell also develops by mitotic cell divisions into the endosperm, which provides the developing embryo with nutrition. Finally the ovule matures into a seed with its coat (testa), which protects the embryo until germination, and the ovary matures into a fruit.

On germination the embryo grows by mitotic cell divisions and cell differentiation into a new plant with roots, stems, leaves and flowers with anthers and ovules. Pictures of the different kinds of cells that comprise a plant can be found in the book by Blackmore (2012). Plant growth is, however, restricted to perpetually embryonic regions called meristems, which can be divided into four types: apical meristems, axillary buds, lateral meristems and intercalary meristems (Moore et al. 1998). Localized growth in meristems means that plants are a mixture of young dividing cells, maturing (enlargement) cells and mature cells. Apical meristems consist of meristematic initials which divide occasionally to provide cells for the rest of the meristematic region. If a spontaneous or induced mutation occurs in a cell capable of division, the outcome will be a chimera; that is an individual composed of tissues of two genetic types whose structure will be determined by the fate of the mutated cell lineage. We will need to consider this topic in detail when we come to consider mutagenesis and mutation breeding in Chap. 16. Here a brief introduction on the origin, structure and properties of chimeras will suffice. It is based on the chapters on chimeras in the book *Plant Mutation Breeding and Biotechnology* edited by Shu et al. (2012).

## Cell Lineages and Chimeras

The shoot apical meristem has a layered structure and a highly ordered orientation of cell division planes, as first described by Schmidt (1924) in his tunica-corpus theory. A typical angiosperm shoot tip has a core tissue (the corpus) covered by two tunica cell layers. They are designated L1, L2 and L3 in species, or stages of development, with three layers, starting from the outermost layer so that L3 is the corpus. The layers are often called histogenic layers. Relative to the meristem surface, the orientation of cell division planes is anticlinal (perpendicular) in the tunica (L1 and L2), but both anticlinal and periclinal (parallel) in the corpus (L3). Cells of L2 origin may divide periclinally after they have left the shoot apical meristem. The direction of cell division is at right angles to its plane of division. Cells of L1 origin give rise to the epidermis, those of L2 origin to the outer mesophyll, the outermost cortical parenchyma and the gametes, and those of L3 origin to the inner cortex, vascular tissues, roots and pith. The arrangement of histogenic layers in the shoot apical meristem is usually transferred to lateral bud meristems. There are different types of chimeras, but the stable ones are usually periclinal chimeras which develop when a lateral bud originates from within the sector bearing the mutated tissue layer. In periclinal chimeras, mutated cells occupy one or more entire histogenic layer(s), with ectochimeras (L1 is mutant layer) the most stable, but often unrecognized because they are phenotypically uniform. Many chimeric cultivars of ornamentals display variegated foliage with periclinal patterns (Marcotrigiano 1997). They are vegetatively propagated through tip or node cuttings which should ensure that their structure remains unchanged, but propagators must regularly remove off-types that have lost the desired pattern. However, in

mutation breeding the desired outcome is usually a homohistic (solid) mutant in which the whole plant (all three layers) has the mutant genotype. This means that the chimeric structure of plants obtained by mutagenesis (called M<sub>1</sub>) needs to be dissociated. The only way to achieve this with certainty is by propagation from a single mutated cell, or groups of cells derived from such a cell. In seed-propagated crops this is usually done by self-pollination of M<sub>1</sub> plants and identification of mutated plants in the M<sub>2</sub> progeny; in other words, through unicellular gametes derived from the L<sub>2</sub> layer. In vegetatively propagated crops use is made of plant regeneration via somatic embryos, adventitious buds from leaf or root cuttings and explants cultivated *in vitro*. If the mutagenic treatment can be applied to materials devoid of pre-existing meristems, followed by regeneration of adventitious buds from single mutated cells, chimera formation can largely be avoided; an example being the use of cell suspension cultures. Finally a word of caution is required. If *in vitro* culture procedures are applied as a means of either chimera dissociation or avoidance of chimera formation, the risk and consequences of other genetic changes occurring (somaclonal variation) need to be evaluated.

## Seeds and Fruits

Given the economic importance of seeds and fruits, some further comments on their structure would seem in order. In a mature seed, the axis of the embryo is divided by the attachment point of the one or two specialized seed leaves (cotyledons). The region above the attachment point is the epicotyl, an embryonic shoot consisting chiefly of a pair of folded miniature leaves enclosing a growing point. This is the first bud of the embryo and is called the plumule. The region below the attachment point is the hypocotyl. An embryonic root (radicle) is often distinguishable at the tip of the hypocotyl. In maize and other cereals, the embryos are partially enclosed in protective sheaths; the coleoptile around the shoot and the coleorhiza around the radicle.

In some seeds with two cotyledons, such as the castor bean, the embryo is surrounded by the nutrient endosperm whereas in others, such as the garden bean, the endosperm is absorbed by the cotyledons so that they serve as the food storage organ. In cereals the situation is more complicated. The grain (also called the kernel) is a one seeded fruit with the seed coat (testa) completely fused to the wall of the fruit. Just inside the seed coat is the aleurone layer. A wheat grain comprises 15 % skin (pericarp, testa, aleurone layer), 82 % endosperm (which is the flour), and 3 % embryo. The outside of the grain is referred to as bran. Furthermore, in cereals the single cotyledon is so highly modified for absorption of nutrients from the endosperm that it has a special name, the scutellum. All seeds store plant food in varying proportions of carbohydrates, fats and proteins, and hence they are a rich source of food for animals and humans, as well as fats and oils for industrial purposes.

The mature ovary, with or without seeds, becomes the fruit. The majority of flowering plants have fruits composed of a single ovary and are referred to as simple fruits. In a mature fruit the ovary wall, called the pericarp, and comprising exocarp,

mesocarp and endocarp, may be fleshy or dry. Fleshy fruits may be berries (tomato, grape, pepper, muskmelon and citrus), drupes or stones (peach, cherry, plum and olive), or pomes (apple, pear and quince). Dry fruits can be divided further into dehiscent and non-dehiscent ones. Examples of dehiscent fruits are legumes and pods (pea and bean), follicles (magnolia and milk weed), capsules (poppies, irises, lilies and jimson weed) and siliques (crucifers such as mustard and radish). Examples of non-dehiscent fruits are achenes (sunflower and buttercup), caryopses (maize and other cereal grains), samaras (maple, ash and elm), schizocarps (carrot, parsley and dill) and nuts (acorn, chestnut, hickory and walnut). In cereals what is usually called the seed or grain is really the fruit or caryopsis. In barley the caryopsis can be naked or hulled, the latter occurring when the lemma and palea adhere to it, these being the bracts (outside parts) of the floret (flower). Aggregate fruits are derived from a flower having many pistils on a common receptacle such as blackberries and strawberries. In contrast, multiple fruits are derived from many separate but closely clustered flowers. Familiar examples of multiple fruits are the pineapple, fig and mulberry, and the beet "seed" is also a multiple fruit. A major achievement of sugar beet breeders in the twentieth century was the reduction of clusters of two to four sessile (stalk-less) blossoms to a single flower with a single ovule, and hence a single (monogerm) seed. This removed the very labour intensive and costly task of thinning young plants by hand.

### ***Six-Rowed and Two-Rowed Barley***

Before leaving seeds and fruits it is worth pointing out how the difference between six-rowed and two-rowed barley arises. The floral structure of the barley plant is a spike with its central axis, the rachis, composed of nodes and internodes. Attached to the rachis nodes are the spikelets, each one comprising an individual floret with surrounding bracts, the lemma and palea, and two subtending outer glumes. Both six-row and two-row barley have three spikelets at each rachis node, but in the latter, only the central spikelet is fertile and will develop into a kernel. Hence, when viewed from above, six-rowed barley appears to have six rows of kernels whereas two-rowed barley appears to have two rows; but that doesn't mean that the latter is lower yielding!

## **Wood**

Although this book is concentrating on food crops, given the economic importance of wood, some further comments on its structure would seem in order; thus providing an example of differences in cellular differentiation. The temperate coniferous forests together with those of the taiga are the world's main source of commercial soft wood timber. This is used in the construction and furniture industries and also converted into woodchip for hardboard or pulp for the paper

industry. The wood anatomy of some gymnosperms is soft and fibrous making it unsuitable for commercial use. It comprises wide rays of parenchyma cells that radiate out through the trunk and make up the bulk of the wood. Although these are still present in the timber producing conifers, they have a harder wood most of which is xylem tracheids. Commercially, however, it is referred to as soft wood to distinguish it from the even denser hard wood timber of many angiosperm trees. A characteristic of most angiosperms is their xylem vessels, which unlike tracheids allow the flow of water directly through interconnected cylinders rather than diffusion across cell membranes (Blackmore 2012).

## Cytological Techniques for Studying Chromosomes

It is now time to move from cell differentiation to cell division. Detailed accounts and excellent black and white photographs of mitosis and meiosis can be found in textbooks written in the 1960s and 1970s, such as *Chromosome Marker* (Lewis and John 1963) and *The Matter of Mendelian Heredity* (Lewis and John 1972). They explain the cytological techniques used to make and study chromosome preparations, the most common of which was the squash method of preparing microscope slides. This involved pre-treatment of the plant material, such as the use of colchicine to aid the spread of chromosomes in mitosis; fixation in acetic-alcohol mixtures to preserve the likeness of the living cell; storage, if necessary, in a refrigerator in 70–95 % alcohol, for example; and finally staining with acetic stains, squashing with the flat end of a brass rod, mounting a cover-slip, and making the slide permanent. Anthers can be used to study meiosis in pollen mother cells and mitosis in pollen grains. Female meiosis is more difficult to study. Mitosis can also be studied in root tips, but with Feulgen stains which require a more complicated procedure involving hydrolysis. In a good preparation, the chromosomes are stained deeply and the cytoplasm hardly at all. When differential staining occurs (banding), chromosome identification is fast and reliable; for example, Giemsa staining of condensed (metaphase) chromosomes was developed in the 1970s (Schweizer 1973). A-T base-pair rich, gene poor, heterochromatin shows up as dark bands. The whole array of chromosomes with their characteristic sizes and centromere positions is called the karyotype of the organism. Karyotyping of species with small and numerous or morphologically similar chromosomes, such as potato, proved difficult by these ‘conventional’ methods and had to await the development of molecular cytological techniques (Gavrilenko 2011).

### **FISH and GISH**

Since the end of the 1980s, molecular cytogenetic techniques have been developed which involve hybridization of DNA probes onto target DNA, usually chromosome

(pachytene) preparations on a microscope slide (Devi et al. 2005; Chester et al. 2010). If the DNA probe is labelled with a fluorescent dye, or a conjugate that reacts with a dye, then any hybridization can be visualized under a fluorescence microscope. Fluorescence *in situ* hybridization (FISH) is used to map defined DNA sequences (including genes) physically and precisely to a specific region of a chromosome. FISH works well with unique species-specific low-copy-number or repetitive sequences (e.g. 5S and 45S ribosomal DNA) and with genomic clones (e.g. bacterial artificial chromosomes) as probes. Genomic *in situ* hybridization (GISH) uses whole genomic DNA from one species to detect its presence in a hybrid with another species, whose unlabelled genomic DNA is added in excess concentrations to block any unwanted hybridization with the probe. It has proved an excellent technique for differentiating the genomes of parental species in natural and artificial hybrids. As a consequence, GISH commonly replaces conventional meiotic pairing analysis of polyploids. Multicolour GISH is a modification which uses mixtures of differently labelled genomic DNA of both parental species and simultaneous hybridization of both probes with chromosomes of allopolyploids (Gavrilenko 2011). It has the potential to discriminate genomes of species that have relatively close affinities to each other; that is greater than 80 % sequence homology, such as wild tetraploid potato species of the series Longipedicellata (Gavrilenko 2011). Together, FISH and GISH provide powerful techniques to confirm the presence and chromosomal location of defined DNA sequences in plant material, including chromatin (chemical content of chromosomes) transferred from wild to cultivated species. In potato FISH techniques have allowed chromosome identification and integration of chromosomal and genetic maps, comparative mapping of potato and related species (e.g. tomato), revelation of the number and chromosomal distribution of tandem repeats such as intergenic spacers (IGS) and interstitial telomeric repeats (ITR), and physical mapping to chromosome VIII of a major late blight resistance gene from *S. bulbocastanum* (Gavrilenko 2011). More examples with further details of these techniques will be given later in this book.

## Mitosis

Mitosis can be followed in living cells by phase-contrast photography, such as Bajer's ciné films of mitosis in the endosperm of the lilies *Haemanthus katherinae* and *Clivia cyrtanthiflora* (Bajer 1957–1958). The stages (in parentheses) of normal mitosis in flowering plants can be briefly described as follows, although there are rare exceptions to the details, which are mentioned in textbooks of botany and genetics. At the start of cell division, the chromosomes gradually become more distinct (prophase) as they shorten and thicken, and are seen as double structures, each having divided longitudinally into identical halves called sister chromatids. The nuclear membrane (envelope) breaks down, thus releasing the chromosomes into the cytoplasm of the cell. During the next stage (metaphase), a non-staining spindle-shaped structure forms in the cytoplasm and is usually attenuated towards

its ends (the poles). The chromosomes become arranged in the equatorial plane across the middle of the spindle. Each chromosome is usually seen bent at a sharp angle whose tip coincides with a non-staining constriction called the centromere, which attaches to a spindle-fibre. The sister half-centromeres then start to separate (anaphase), and to move in opposite directions towards the poles. Finally (telophase), nuclear membranes form round each polar group of chromatids, which gradually become longer and thinner until they can no longer be clearly seen (interphase). The completion of division of the cell into two new cells, each with a completely enclosing cell membrane, and equipped with a nucleus containing the same set of chromosomes as the mother cell, usually follows closely after telophase. Thus the cell cycle comprises cell growth (interphase), then division of the cell nucleus (mitosis), and finally division of the cytoplasm (cytokinesis) to give two daughter cells each with a nucleus.

Reliable information about the relative DNA content of different nuclei can be obtained by measuring the intensity of pigmentation after Feulgen staining. However, Howard and Pelc (1951) used autoradiography to demonstrate that radioactive phosphorus ( $^{32}\text{P}$ ) was incorporated into nuclei in the meristematic region of *Vicia faba* (broad bean) root-tips during interphase, but not during mitosis. The doubling of the DNA over a comparatively long period of time during interphase was confirmed in other organisms by measuring the absorption of UV light (Whitehouse 1973). Thus the cell cycle is divided into the period of DNA synthesis called the S-phase, the interval between this and mitosis (G2), and the interval between mitosis and the next round of DNA synthesis (G1).

Watson and Crick (1953b) proposed a mechanism of DNA replication in which each new DNA molecule would consist of one old nucleotide chain (strand) and one new one, complementary to the old. One of the first pieces of evidence for this semi-conservative replication of DNA was obtained by Taylor et al. (1957) with root-tips of *Vicia faba* labelled with tritiated ( $^3\text{H}$ ) thymidine (Whitehouse 1973). DNA synthesized in one round of replication was made radioactive by supplying the cells with  $^3\text{H}$ -thymidine. Then the distribution of the radioactive label was followed through subsequent rounds of replication in the absence of the radioactive precursor. This was done by attaching radiosensitive film to microscope slides carrying spread metaphases and noting which chromatids were radioactive as shown by silver grains in the film above them. The results demonstrated that the chromosome replicates semi-conservatively and were compatible with semi-conservative DNA replication. This was subsequently confirmed in other organisms and since then research has revealed the molecular details of the process. DNA synthesis occurs at multiple replication origins along the chromosomes, in loops of separated strands, on both strands in the 3' to 5' direction along the strand. In other words the new strands are synthesized in the 5' to 3' direction. However, for our purposes, the outcome but not the details are important. Each new daughter cell contains the same genetic material, and this is also true for the regrowth of individuals and for new individuals produced vegetatively or asexually, whether naturally or through human intervention.

## Meiosis

The stages (in parentheses) of normal meiosis (divisions I and II) in flowering plants can be briefly described as follows, with prophase I subdivided into its own five stages. The chromosomes first become visible as exceedingly fine single threads (leptotene). Structurally identical, homologous chromosomes then start to pair (zygotene), shorten and thicken, so that eventually they are paired throughout their length in the form of bivalents (pachytene). The technical term for this chromosome pairing is synapsis. Then a remarkable change in appearance takes place. The homologous members of each bivalent begin to separate, but do not dissociate. In the regions of separation each homologue is divided longitudinally into two chromatids, but at one or more locations along the length of the bivalent, one chromatid from each homologue exchanges its partner. Each cytologically visible crossover is known as a chiasma. As the separation of homologues proceeds, loops form between consecutive chiasmata while half loops are produced between a chromosome end and its nearest chiasma. Initially the loops are all in one plane (diplotene), but then successive loops lie at right angles to one another (diakinesis). Separation may force the chiasmata to move towards the chromosome ends, a movement known as terminalization. The shape of the bivalent is thus determined by the length of the chromosome, the position of the centromere and the number, distribution and degree of movement of the chiasmata.

As in mitosis, the disruption of the nuclear membrane signals the development of a spindle (metaphase I). Each bivalent attaches itself to the spindle by its two centromeres in positions equidistant above and below the spindle equator. There is a lapse of the intimate association between sister chromatids which allows the chiasmata to slip completely apart and enables the half-bivalents to move apart towards opposite poles of the spindle under the influence of their respective centromeres (anaphase I). The technical term for this separation to opposite poles is disjunction. Each bivalent contributes one chromosome to each polar group, thus overall each group comprises a complete set of homologous chromosomes, with each chromosome consisting of two chromatids. Nuclear membranes may develop round each group of chromosomes (telophase I). The cytoplasm too may divide and even a cell wall may form. Alternatively the second division of meiosis can follow immediately without a brief interphase in which the chromosomes become difficult to observe. Either way, the chromosomes are short and thick after prophase II.

Mechanically the second division of meiosis is similar to a mitotic division, although the homologous chromatids are widely separated from one another except in the undivided regions at or near the centromere which serve to hold them together. The nuclear membranes break down again and two new spindles form (metaphase II). Where division occurs in a common cytoplasm or in two associated though distinct cytoplasmic areas, the relative orientation of the two spindles often varies both within and between species. In some cases however, second division spindles have a consistent arrangement, as in the linear tetrads on the female side in flowering plants and in the Ascomycete fungi. As in mitosis, the centromeres of all

of the chromosomes align themselves on the equator of the second division spindle. The sister-chromatids of each chromosome then separate and move towards opposite poles on each spindle (anaphase II), followed by the formation of nuclear membranes round each of the four groups of chromatids (telophase II). Thus four nuclei are produced by each meiotic sequence and each nucleus contains a complete set of single chromatids. As already mentioned, all four products develop into gametes on the male side, whereas only one megasporangium is usually functional, even when two or all four products are involved in the production of the mature embryo sac.

## Key Interpretations of Meiosis for Heredity

In the early 1900s, in organisms where all of the chromosomes could be individually recognized, those which associated in pairs were shown to be identical in length and morphology (homologous) and of maternal (egg) and paternal (sperm) origin, respectively (Whitehouse 1973). Furthermore, there was usually random orientation of different chromosome-pairs on the first-division spindle, as observed in rare examples of differences in homology between paired-chromosomes. In contrast, the origin of the four strands (chromatids) which make up the diplotene and diakinetically aligned chromosome pairs was a source of controversy for many years. Janssens (1909) put forward the hypothesis that chiasmata are invariably the consequence of previous crossing-over between two of the four chromatids; in other words, paternal and maternal chromatids made contact at intervals, then broke and rejoined paternal to maternal and vice versa. The other two chromatids, one paternal and one maternal, each remained intact. Janssens thought that the exchange took place during diplotene. Later, after studying the meiotic chromosomes in various flowering plants, Belling (1928) suggested that the exchange took place at the preceding pachytene stage, while the homologous chromosomes were closely associated, despite the fact that chromatids and exchanges could not be directly observed. This modified hypothesis proved correct. Cytological evidence accumulated from studying unequal chromosome pairs, polyploids and chromosomal inversions (Whitehouse 1973), but final proof came from isotope labelling of chromosomes (Taylor 1965; Peacock 1970; Jones 1971), in animals rather than plants for ease of experimentation.

At meiosis, like mitosis, the replication of DNA takes place in interphase before the nuclear division begins. This was demonstrated in plant material by Taylor (1953) using autoradiography. Following incorporation of radiophosphorus ( $^{32}\text{P}$ ) into flower-buds of *Lilium longiflorum* and *Tradescantia paludosa*, he found that the period of DNA synthesis in the pollen-mother-cells differed, but was completed well before the zygotene stage when homologous chromosomes began to associate. Later he (Taylor 1965) studied meiosis in males of the grasshopper *Romalea microptera*. He introduced  $^3\text{H}$ -thymidine at the penultimate DNA replication before meiosis so that at meiosis half the chromatids were labelled and half were not.

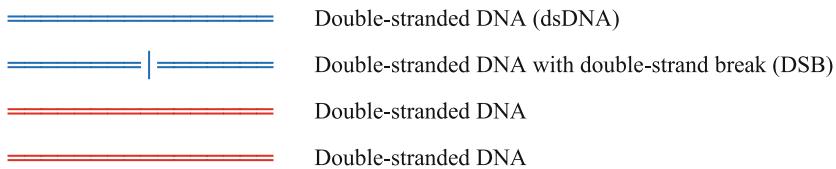
It was then possible to observe that chiasma frequency and distribution at diplotene corresponded to the frequency and distribution of exchanges between labelled and unlabelled chromatids, taking account of the expectation that only half the crossovers would be between a labelled and an unlabelled chromatid. Since the 1960s, models have been put forward of breakage and rejoining of chromatids through hybrid DNA formation, and progress has been made in understanding the biochemistry of the process.

Acceptance of the correctness of the Janssens-Belling theory of chromatid exchange meant that inferences about meiosis could be made with confidence from genetic experiments, and causal connections could be made between cytological and genetic associations, as we shall see in the next chapter. The ideal situation in which to obtain such information is where the four products of a single meiosis can be recovered in isolation from those of neighbouring meioses. This tetrad analysis is possible in some plants and proved of tremendous value for genetic research in Ascomycete fungi such as *Neurospora crassa* and *Sordaria fimicola* (Lindegren 1933; Whitehouse 1973).

## Molecular Basis of Genetic Recombination

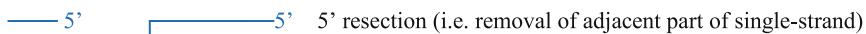
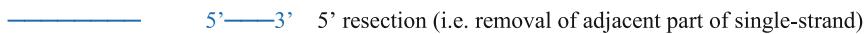
The molecular basis of genetic recombination during meiosis involves the repair of double strand breaks. It has therefore been suggested that damage to the genome may have provided selection pressure for the evolution of genetic recombination, initially as a means of exchanging ‘good’ bits of DNA for ‘bad’ (Friedberg et al. 2006). Our current understanding of the molecular basis of meiotic recombination in higher plants has come from extensions of studies in *Saccharomyces cerevisiae* (budding yeast) to *Arabidopsis* and other plants, such as rice and maize. The detailed review by Osman et al. (2011) explains why the meiotic recombination pathway in *Arabidopsis* is thought to be broadly similar to the ‘early crossover decision’ model which has been developed to explain recombination in budding yeast. Knowledge of the genes and protein activity required for crossover formation during plant meiosis has come from studying many mutants and their phenotypes. Homologous recombination is initiated during leptotene by the formation of programmed double-strand breaks (DSBs) in the DNA of a chromatid, catalysed by the conserved protein called Spo11. The process is complicated because the breaks are ‘repaired’ by DNA synthesis using single stranded DNA (ssDNA) as template. The process is complete by the end of pachytene. Figure 3.3 is a very simplified diagram of the start, key steps, and possible end products of the process, but will suffice for our purpose. In three dimensions the sister chromatids are joined together along their length by a proteinaceous axis at the bottom of their loops, but for simplicity the diagram represents them as a straight line. Close pairing (synapsis) of homologous chromosomes is achieved through the formation of the synaptonemal complex which includes a transverse filament protein, but the details need not concern us. The review by Osman et al. (2011) has coloured pictures and

**Start:** there is cohesion between sister-chromatids (same colour) and pairing between homologous chromosomes (different colours).



We will focus on centre pair of non-sister chromatids.

**Next-steps:**

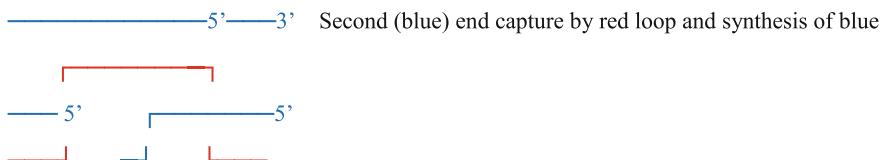


**Finish:**

**a**



Step required for homologous recombination:



**b**



Black represents one blue strand and one red strand in DNA (heteroduplex DNA)

**Fig. 3.3** Homologous recombination. (a) Non-crossover (NCO) as a result of synthesis-dependent strand annealing (b) crossover (CO) and (c) non-crossover (NCO) resolution of double 'Holliday' junction

explains how features can be visualized on chromosome spreads through immunolocalization using antibodies that recognize components.

After break formation, Spo11 is covalently attached to the 5' ends of the DNA on either side of the break site. It is removed as the DNA ends are resected to generate 3'-ended single stranded DNA (ssDNA). These single strands interact with two homologues (Rad51 and Dmc1) of the protein RecA to form nucleoprotein filaments. One filament then invades the homologous duplex (double stranded) DNA of a non-sister chromatid to form a stable single-end invasion intermediate. There is evidence in budding yeast for this being a key control step in determining the outcome of DSB repair. In *Arabidopsis*, meiotic recombination is initiated by approximately 150 DSBs, about 10 of which are repaired as crossovers (COs) and 140 as non-crossovers (NCOs). The displaced DNA strand forms what is called a D-loop (displacement loop) and this extends as the invading strand polymerizes (DNA synthesis on single strand template). Crossovers then result from the following process known as double strand break repair (DSBR). The displaced DNA strand captures the 3' end on the other side of the break and acts as template for DNA synthesis. Subsequent ligation (joining) of the broken DNA strands results in the formation of the double Holliday junction (dHJ) recombination intermediate (named after the proposer of the model) in which there is an exchange of strands. It is not yet known in *Arabidopsis* if resolution (processing) of these junctions gives only crossovers (Fig. 3.3b), or if a proportion result in non-crossovers (Fig. 3.3c). However, it is proposed that most, if not all, non-crossovers arise from synthesis-dependent strand annealing in which there is no capture of the second end by the D-loop (Fig. 3.3a). In synthesis-dependent strand annealing the blue 3' end is extended, rejected from red, annealed to blue and the process completed by synthesis and ligation of both blue ends to their blue strands, and hence no crossover.

In most species, accurate chromosome disjunction during meiosis requires at least one 'obligate' crossover per chromosome pair. As we will see in the next chapter, crossovers in general are non-randomly distributed and multiple crossovers on the same chromosome are spaced apart; in other words, crossover interference occurs. Hence DBS sites must be non-randomly selected to mature into crossovers, and this has led to the 'early decision' model of meiotic recombination. Interestingly, there is evidence for a second meiotic recombination pathway not subjected to crossover interference (Osman et al. 2011). Clearly there is still much to learn about the molecular basis of meiosis, but with so many genes and protein products involved, there should be the prospect of manipulation by plant breeders. In cereals and forage grasses, for example, the distribution of crossovers is heavily skewed to the ends of the chromosomes. These end regions are gene rich, but a significant proportion of genes reside where there is little recombination. Hence increasing recombination frequency or changing crossover distribution could provide breeders with more genetic variation to exploit (Osman et al. 2011).

## Changes in Chromosome Number: Errors in Distribution

Chromosome number is a characteristic of a species and a reflection of the accuracy of chromosome movement and distribution during the two divisions of meiosis. But low levels of error do occur and result in changes of chromosome number; and sometimes these changes are perpetuated by natural or human selection. Variations in chromosome number produce two types of individuals or cells: those whose somatic complements are exact multiples of the basic haploid number characteristic of the species (euploid); and those with irregular chromosome numbers (aneuploid). Euploids will be dealt with in later chapters: we will consider unreduced gametes and polyploids in Chap. 8 and the production of haploids and doubled haploids in Chap. 13. Here we will concentrate on aneuploids. An individual with an extra chromosome is designated  $2n + 1$  (trisomic), and one with a loss of a chromosome as  $2n - 1$  (monosomic). A variety of mechanisms can result in cells deficient or duplicated for a particular chromosome. Details can be found in classical textbooks on cytogenetics such as *Chromosome Marker* (Lewis and John 1963). Nondisjunction is a recognized source of aneuploid cells. At mitosis this leads to both sister chromatids passing to the same pole and so producing a complementary pair of trisomic-monosomic daughter cells. At meiosis, failure of a pair of homologous chromosomes to pair and cross over results in two univalents at metaphase. These do not orient properly between the poles and may fail to segregate, or may pass randomly to one pole or the other, or may divide into their two chromatids and hence be unable to divide and segregate properly during the second meiotic division. Fusion of an  $n + 1$  and an  $n - 1$  gamete with a normal  $n$  gamete results in a  $2n + 1$  and a  $2n - 1$  individual, respectively. Unbalanced gametes are also produced from triploids where three homologous chromosomes 'pair' to give a bivalent plus univalent. Hence trisomics can be produced from crossing a diploid with a triploid individual. In *Datura stramonium* ( $2n = 2x = 24$ ) all 12 different trisomics have been identified and shown to have characteristic phenotypes; for example, they differ in seed pod morphology (Swanson et al. 1967). The average frequency of transmission of  $n + 1$  gametes through the egg is about 20 %, but varies with chromosome, whereas most of the trisomics in *Datura* are not transmitted through the pollen (Swanson et al. 1967). Monosomics are rare because the loss of a chromosome can expose recessive lethals. Another source of variation in chromosome number is centric fusion. Two acrocentric (centromere at end) chromosomes can fuse to form one chromosome with a more or less medially placed centromere, and the reverse process can occur. The details need not concern us here but can be found in the book by Lewis and John (1963). Strictly speaking, centric fusion is a special category of unequal interchange between two acrocentric chromosomes.

Supernumerary (B-) chromosomes occur in some plants (e.g. some strains of maize) and are largely composed of heterochromatin together with a centromere. As a consequence they have little or no detectable effect on phenotype, including the viability of the plant. They will not be considered further here, but examples can again be found in the book by Lewis and John (1963).

## Importance of Meiosis in Plant Breeding

The outcomes of meioses determine the combinations of genes that are passed on to the next generation through sperm cells and egg cells and hence the results of hybridization experiments, whether for breeding or genetic studies. The outcomes also determine the size of breeding programme and number of generations required to achieve specific objectives and hence the rate of progress. Recognition of desirable genes (alleles), phenotypically, genotypically or through linkage to markers, can ensure that the correct products of meioses are selected in a breeding programme, but cannot influence the creation of those products. Hence recognition can ensure certainty of outcome in breeding programmes but does not necessarily reduce the time required to breed a new cultivar, nor necessarily make the process cheaper and more efficient. The same is true of genes discovered from sequencing plant genomes, and then tracked through hybridizations in a breeding programme rather than cloned and used for genetic transformation. Given the importance of the outcomes of meioses in breeding, it is not surprising that there is interest in meiotic mutants and their potential role in plant breeding. For example, Wijnker et al. (2014) have developed a protocol known as reverse breeding whereby parental lines can be generated from uncharacterized heterozygotes, so that any selected heterozygotes can be recreated as F<sub>1</sub> hybrids. The protocol uses a meiotic mutant to suppress meiotic crossovers in a hybrid plant (achiasmatic meiosis) to ensure the transmission of non-recombinant chromosomes to haploid gametes. These gametes are subsequently regenerated as doubled-haploid (DH) offspring. The protocol works in *Arabidopsis* and now needs to be transferred to crop plants. We will finish this chapter with a brief consideration of the origin of genetic variants.

## DNA Damage

The nature of the external environments in which plants grow and the internal environments of plant cells, means that their DNA can be altered by chemicals through hydrolysis, deamination, alkylation and oxidation, and by both ionizing and ultraviolet (UV) radiation. The outcomes of DNA alteration can be mismatched and chemically modified bases along the double helix and both single and double strand breaks to the helix. These can alter the integrity, replication and transcription (expression) of the genetic information with a range of consequences, some more serious than others. Double-stranded DNA breaks are potentially the most serious type of damage. Accumulated damage can preclude cell division and eventually kill non dividing cells through its effects on gene expression and therefore plant metabolism. Not surprisingly, all living organisms, including plants, have evolved mechanisms to repair (and also to tolerate) this damage, but not completely as genetic variants (mutations) are the raw materials of evolution. In other words,

evolution has struck a balance between damage and repair, but one that can be altered. Likewise, we can assume that a similar balance has been struck in the error rate (incorporation of ‘wrong’ bases) that occurs during DNA replication, where it is known that DNA polymerase also acts as a proof reader. Furthermore, various types of DNA damage and arrested DNA replication are known to activate specific cell cycle checkpoints that result in arrested cell cycle progression, thus providing more time for repair or damage tolerance to occur (Friedberg et al. 2006). A final point (Britt 1996) is relevant to seed storage and quality as well as to mutation breeding. Although the exact structure and hydration state of DNA in dried seeds is unknown, DNA repair does not occur and one would therefore expect an accumulation of spontaneous mutations. At some point the amount of damage incurred as a consequence in aging seeds may exceed the repair capacity of the germinating seedling.

In Chap. 16 we will explore how plant breeders have deliberately increased the exposure of plants to various chemicals and radiation in order to induce a higher rate of mutation, followed by a search for desirable mutants among the wreckage (mutation breeding). For example, ethyl methane sulfonate (EMS), an ethylating agent, is widely used to induce point mutations and ionizing radiation to induce chromosomal breaks, inversions, duplications, and translocations. Here, however, we are going to consider repair mechanisms in order to understand when mutations are an outcome.

## DNA Repair

DNA repair mechanisms have been reviewed by Britt (1996), Waterworth et al. (2011), Mannuss et al. (2012) and Shu et al. (2012). Brief summaries of the main repair mechanisms follow. Tolerance is where DNA lesions are in effect by-passed rather than repaired during replication; but this increases the risk of mutagenic mismatches or rearrangements. Details can be found in Shu et al. (2012).

## Photoreactivation

Photoreactivation is where the biological effects of UV-B radiation (wavelength 280–315 nm) are significantly reduced by subsequent exposure to light in the blue or UV-A range (315–400 nm) of the spectrum. The primary effect of UV-B radiation on DNA is the formation of pyrimidine dimers from cross-linking pairs of pyrimidine (thymine and cytosine) bases occupying adjacent positions in the DNA chain. Every dimer acts as a block to transcription and replication, while only a small fraction of dimers results in a mutation. As effects can be marked in epidermal cells, plants are thought to produce natural sunscreens such as flavonoid pigments which can absorb photons in the UV-B and UV-A range. Photoreactivation is mediated by

the action of photolyase enzymes which bind to the dimers, and upon absorption of a photon of the appropriate wavelength (350–450 nm), directly reverse the damage in an error-free manner. There is evidence that the repair capacity of plants depends on the quality and timing, as well as the quantity, of light. Dimers can also be removed by nucleotide excision repair where many different kinds of mutation result from error prone gap-filling, albeit at a low frequency.

## ***Excision Repair***

In excision repair, damaged DNA is replaced with new, undamaged nucleotides. Excision repair pathways fall into two main categories, base excision repair (BER) and nucleotide excision repair (NER). In base excision repair, a single damaged base is recognized and removed through the action of one of many lesion-specific glycosylases, leaving the sugar-phosphate backbone intact. The resulting abasic sites are then recognized by an apurinic/apyrimidinic (AP) endonuclease or AP lyase, which nicks the backbone of the DNA at the AP site. The nicked DNA is then restored to its original sequence through the combined actions of a number of enzymes. Plants have many glycosylases for excising bases modified by alkylation.

Nucleotide excision repair differs from BER in two ways: the remarkably wide spectrum of DNA damage products that are recognized; and initiation of damage removal by generating nicks on the damaged strand. These nicks occur at a specific distance both 5' and 3' from the lesion, which is excised as an oligonucleotide through the action of a helicase. Then the undamaged strand is used as a template to fill the resulting gap so that the repair pathway is essentially, but not completely, error free.

Finally, mismatch repair (MMR) specifically concerns the removal of mispaired bases in DNA (Friedberg et al. 2006). Mismatches are mutagenic lesions that occur during the replication of both undamaged (frequency  $10^{-7}$  to  $10^{-8}$ ) and damaged ( $10^{-2}$  to  $10^{-3}$ ) DNA. Mismatch repair involves recognition of mismatch, incision and removal of nascent strand around mismatch by an exonuclease, and re-synthesis and rejoicing of oligonucleotide to fill gap. It decreases error rates by 10- to 100-fold.

## ***Double Strand Break Repair by Non Homologous End Joining***

DNA sequences near the ends of double-strand breaks (DSBs) are rapidly degraded, with the consequence that they expand into gaps that cannot simply be religated to restore the original sequence. They are, however, rejoined end to end in what appears to be a random fashion. The process is sometimes referred to as ‘illegitimate recombination’, but is more commonly called non homologous end joining

(NHEJ). It is thought to be the main method of DSB repair in the vegetative tissues (including meristem initials) of higher plants. NHEJ often produces short deletions and insertions because it involves joining broken ends back together without regard for absolute DNA sequence fidelity. More extensive chromosome rearrangements can also occur. Recent research indicates that at least four different end joining pathways operate in plants but only one is known in detail, and these details can be found in the review by Mannuss et al. (2012). The first step is the recognition and binding of the broken DNA ends by proteins (Ku70-Ku80 heterodimer) which prevent large-scale degradation of the ends and bring them into close proximity. These proteins interact with other proteins (the Mre11, Rad50 and NBS1 complex) involved in DNA end processing and damage signalling. These end processing factors modify the DNA to produce ends suitable for ligation (rejoining) by another enzyme complex (XRCC4/LigIV) with DNA ligase activity. The processing of the DNA ends can result in deletions or insertions and is the reason why NHEJ based repair often results in mutations in the repaired DNA. An alternative end-joining pathway is micro-homology-mediated end joining (MMEJ) (McVey and Lee 2008). It uses 5–25 bp microhomologous sequences for alignment of break ends during re-joining, and results in deletions flanking the original break. Insertions of ‘filler’ DNA also occur, indicative of a DNA synthesis-dependent repair process.

We shall see in Chap. 17 that NHEJ is the mechanism by which DNA is inserted during transformation. Topologically, T-DNA integration represents a DSB repair event, although the mechanism of integration could also occur via single-strand intermediates, as detailed in several possible models. Transgene integration and DSB repair are closely linked processes and numerous studies conducted 10–20 years ago provided much insight into the mechanisms of illegitimate recombination in plants through analysis of transgene integration sites. Despite our increase in knowledge, the mechanisms of transgene integration are still largely unknown. In particular, the insertion mechanism(s) of T-DNA, which is an essential step in the transformation of plants using *Agrobacterium*, remains obscure, despite the importance of this process to biotechnology.

## ***Double Strand Break Repair by Homologous Recombination***

It was mentioned earlier that the NHEJ pathway is thought to be the main mechanism for the repair of double strand breaks. However, if homologous DNA strands (for example, sister chromatids after DNA replication) are present in the cell, then repair by homologous recombination is a possibility. Several alternative pathways may exist in plants, but there is good evidence for repair by synthesis-dependent strand annealing (SDSA) (Waterworth et al. 2011; Mannuss et al. 2012). Furthermore, if a double strand break occurs between two repeated sequences in the same orientation, then single strand annealing (SSA) can be used because the repeats provide sufficient homology for annealing. Adjacent to the break, single-stranded DNAs are created by 5' digestion so that the repeated sequences

can be used as complementary strands to anneal the ends of the break, after which non-homologous tails are detached and the nicks ligated. The genetic information between the two repeats is lost.

## Mutations, the Source of Genetic Variation

Mutations result from errors in DNA replication and from DNA damage that are not corrected by the repair mechanisms considered in previous sections. They are heritable changes in the genetic material and hence the source of genetic variation within species on which both natural and human selection operates. Transposable (genetic) elements can also cause mutations, as we shall see at the end of this chapter. In Chap. 16 we will look at induced mutagenesis and mutation breeding. Here it is worth mentioning that once genomes have been sequenced, one needs to find out the functions of all the genes. One approach is to use mutagenesis to knock-out each gene in turn and determine the effect on phenotype.

With the power of “Next-Generation Sequencing” (NGS) methods, genetic differences can be detected at the DNA level. They may be base-pair substitutions (SNPs); insertions and deletions (InDels), both small (often recognized as frame-shifts) and larger ones recognized by classical cytogenetics; inversions (paracentric and pericentric); and translocations between non homologous chromosomes. As mutations are rare events, they will arise in one of a pair of homologous chromosomes. Despite this, at least in maize, a high level of polymorphism has been found. On average, maize inbred lines B73 and Mo17 contain an InDel every 300 bp and a SNP every 80 bp, and within expressed genes SNPs are found between the inbred lines on average every 300 bp (Springer et al. 2009).

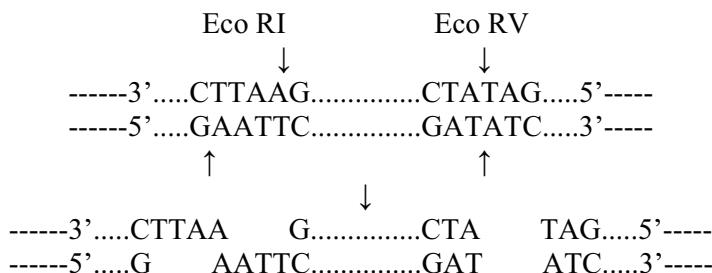
## Base-Pair Substitutions

Base-pair substitutions result in single nucleotide differences between homologous chromosomes, referred to as single nucleotide polymorphisms (SNPs). They can be transitions (purine/pyrimidine to purine/pyrimidine) or transversions (purine/pyrimidine to pyrimidine/purine). These can be used as molecular markers in plant breeding when they occur either in or tightly linked to genes (alleles) of interest for trait improvement, as we shall see in subsequent chapters. In addition to SNPs, four other types of DNA markers have been commonly used in plant breeding and will also be encountered in subsequent chapters (Heldt and Piechulla 2011). The basic concepts and types of marker have been reviewed by Collard et al. (2005) where extensive references to the original literature can be found.

### **RFLP Markers**

Base-pair substitutions, insertions and deletions may create or eliminate palindromic DNA sequences that are recognized by bacterial restriction endonucleases (enzymes), the result being restriction fragment length polymorphisms (RFLPs). The various restriction endonucleases have specific recognition sites of four to eight base pairs, but commonly six. Since these restriction sites occur at random it is possible to cleave the genomic DNA of a plant into thousands of fragments. The restriction endonuclease Eco RI makes staggered cuts whereas Eco RV produces blunt ends.

Restriction fragments are separated by length in an agarose gel using electrophoresis and transferred to a nitrocellulose membrane. The buffer used causes dissociation of the DNA fragments into single strands. A radioactively (or fluorescently) labelled DNA probe is added and hybridizes with complementary DNA sequences on the membrane, which can then be visualized on an X-ray film (or by fluorescence). The probes (RFLP markers) are DNA sequences of interest, usually 10–20 kbp in length. RFLPs are single locus, multi-allelic, codominant markers.



### **Markers Produced by PCR (Variation in DNA Sequence)**

The differences between DNA sequences of individuals or cultivars of a species can be analyzed by the amplification of randomly obtained DNA fragments (random amplified polymorphic DNA or RAPD), using the polymerase chain reaction (PCR). Selected DNA fragments up to 2–3 kbp in length can be amplified. The ends of the fragments need to be defined by two primers that bind to complementary DNA sequences after heat denaturation of the double-stranded DNA. In the first step, the DNA double strands are separated into single strands by heating to about 95 °C. During subsequent cooling the primers hybridize with the DNA single strands. Then at a medium temperature, a DNA polymerase replicates (synthesizes) the DNA between both primers. A DNA polymerase from the thermophilic

bacterium *Thermus aquaticus* (Taq) is used since it is not affected by the heat treatments. The alternating heating and cooling is usually continued for 30–40 cycles, with the amount of DNA doubled each cycle. After the second cycle, DNA fragments of uniform length are amplified, the length determined by both primers. Very small DNA samples can be rapidly amplified in this way.

-----3'....CTACTAGATG.....CATCTAGTAG....5'---- One copy  
 -----5'....GATGATCTAC.....GTAGATCATC....3'----

GATGATCTAC and CATCTAGTAG primers

-----3'....CTACTAGATG.....CATCTAGTAG....5'----  
 GATGATCTAC→  
 ←CATCTAGTAG  
 -----5'....GATGATCTAC.....GTAGATCATC....3'----

↓ PCR

CTACTAGATG.....CATCTAGTAG Millions of copies  
 GATGATCTAC.....GTAGATCATC

In the RAPD technique, genomic DNA and only one oligonucleotide primer are required for the PCR. Typically the primer consists of ten nucleotides. As the probability of an exact match of ten complementary nucleotides on the genomic DNA is low, the primer binds at relatively few sites and only a few sections of the genome are amplified. The amplified single DNA fragments can be separated by gel electrophoresis, stained with ethidium bromide, and detected as fluorescent bands under ultraviolet light. Point mutations which eliminate primer binding sites or form new ones, and deletions and insertions, affect the size and number of PCR products; and DNA polymorphisms can be detected. Changing the primer sequence can generate different DNA fragments so that primers can usually be found that give DNA band differences which correlate with trait differences of interest. Sequence characterized amplified region (SCAR) markers can be derived by sequencing the ends of RAPD fragments to develop longer primers (about 22–24 nucleotides), and likewise sequence tagged site (STS) markers from RFLP markers.

Amplified fragment length polymorphism (AFLP) markers are based on the selective amplification of sets of restriction fragments from genomic DNA. The number of amplified fragments is limited by using two restriction enzymes, one that cuts frequently (four-base pair recognition site) and one that cuts rarely (six-base pair). Doubled-stranded adaptors are ligated to the fragments and PCR

primers (17–21 bases) designed to match the sequence of the adaptors and recognition site. The number of fragments is further limited by including one to three arbitrary chosen nucleotides at the 3' end of the primer. The polymorphisms result from point mutations in the enzyme recognition sites or the nucleotide extensions, which either allow or prevent PCR amplification of a specific DNA fragment. Insertions or deletions between recognition sites can also give rise to AFLPs. The amplified restriction fragments are radioactively or fluorescently labelled and size separated by electrophoresis on high-resolution polyacrylamide gels. AFLPs, like RAPDs, are multilocus, single allelic, dominant markers.

### ***Micro-Satellite Markers (Variation in DNA Repeats)***

Micro-satellite DNAs (or simple sequence repeats: SSRs) comprise sequences of two to five nucleotide pairs, repeated up to 70 times, which are interspersed throughout the genome. The number of repetitions is highly polymorphic and can be detected by PCR, provided the sequences flanking the repeats are known. The amplification products are size separated by electrophoresis and visualized by silver staining or fluorescent dyes. Micro-satellite mutations are caused by an intra-molecular mutation mechanism called DNA replication slippage. The most common mutations are changes of a single repeat unit. Micro-satellites, like RFLPs, are single locus, multi-allelic, codominant markers. Genomic regions that occur between microsatellites (inter simple sequence repeats: ISSR) can be amplified by primers anchored at either the 5' or 3' end of a repeat region.

### ***Diversity Array Technology (DArT) Markers***

Diversity Array Technology (DArT) markers were developed by Jaccoud et al. (2001) as a high-throughput genome analysis method based on microarray technology. As explained by Gupta et al. (2008), the genotyping array is developed through a discovery array from a particular representation of segments of genomic DNA from a pool of individuals (metagenome) representing the genetic diversity of a species. The level of repetitive DNA sequences is reduced through a process called complexity reduction. Individual DNA clones from the genomic representation, made using a suitable vector and *E. coli*, are amplified and spotted onto glass slides to give the desired 'discovery array'. Labelled (fluorescent dyes) genomic representations of the individual genomes included in the metagenome pool, are then hybridized to the discovery array, and the polymorphic clones (called DArT markers) assembled into the 'genotyping array' for routine genotyping work. Genotypes for each DArT marker are scored using hybridization with the genotyping array. Data acquisition is automated through image analysis which compares the intensity of green and red fluorescence. The details need not concern us here, but the

interested reader can easily find them in the literature for particular plant species; for example, the paper by Sansaloni et al. (2010) for *Eucalyptus*. Diversity arrays generally detect polymorphisms due to single base-pair changes (SNPs) at the restriction sites of endonucleases (used to produce segments of genomic DNA), and to InDels/rearrangements within restriction fragments (Jaccoud et al. 2001). The technology is increasing in popularity (Gupta et al. 2008).

## Deletions, Inversions and Translocations

After a deletion, the portion of the chromosome carrying the centromere functions as a genetically deficient chromosome, whereas the rest of the chromosome does not survive as it has no means of movement in anaphase. Because deficiencies result in the loss of genes, they can have serious deleterious effects on an organism, and may be lethal in the homozygous condition. Large heterozygous structural rearrangements of chromosomes may suppress recombination and may result in reduced fertility. Explanations of how these changes can arise and be recognized, and their consequences during meiosis, can be found in classical textbooks on cytogenetics such as *Chromosome Marker* (Lewis and John 1963), whereas newer books give more up to date accounts of *The Role of Chromosomal Change in Plant Evolution* (Levin 2002). We will finish this section with an example of an inversion in a model plant system, a survey of translocations in a crop and chromosomal evolution in a plant family. We will then finish the chapter with a brief account of transposable elements.

## Inversions in the Monkeyflower

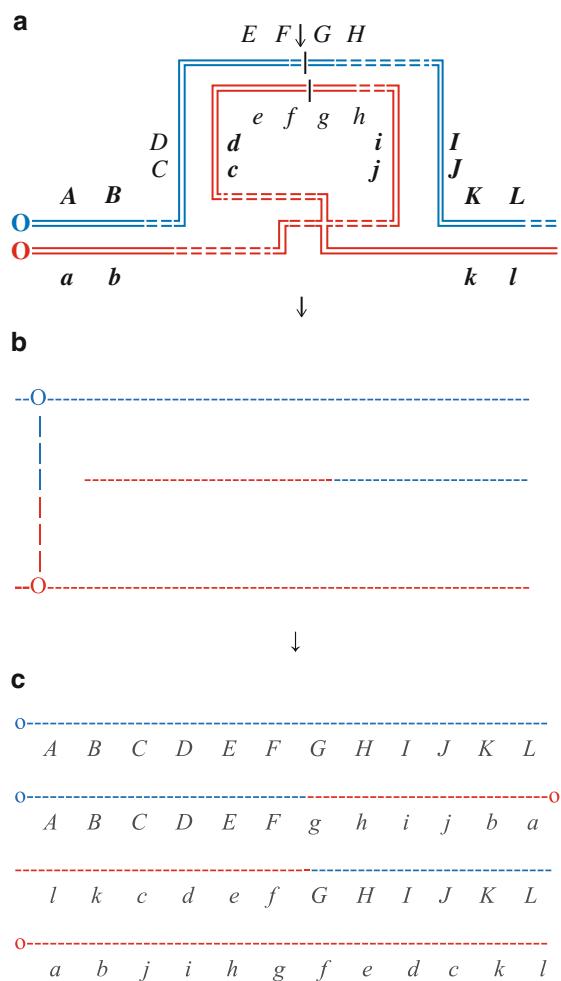
The yellow monkeyflower *Mimulus guttatus* is an outbreeding species that occurs in western North America and has proved an excellent model system for ecological genetics (Zuellig et al. 2014). It can be self-pollinated to produce inbred lines. Two ecotypes have been found, one a facultative annual the other a perennial. Inland annual populations are locally adapted through early flowering to escape the onset of summer seasonal drought. When the more slowly growing perennials are transplanted into inland habitats they die before flowering. In contrast, persistent summer fog and lower temperatures along the Pacific Ocean allow coastal perennial populations to grow to a much larger size before flowering than transplanted annual plants, ultimately resulting in substantially greater seed set in the first year and survival over multiple years. The actual distribution of ecotypes is an overlapping mosaic of discrete annual and perennial populations depending on local environmental conditions, namely reduced soil water availability in the summer versus habitats with high year-round soil moisture. Lowry and Willis (2010) found that a chromosomal inversion polymorphism (on linkage group 8 out of 14) contributed to

the adaptive divergence and reproductive isolation between the annual and perennial ecotypes of *M. guttatus*. They used a novel reciprocal transplant experiment involving outbred lines to avoid inbreeding depression, where alternative arrangements of the inversion were reciprocally introgressed into the genetic backgrounds of each ecotype. The details are worth reading but need not concern us here. The AN arrangement of the inversion was consistently found in annual populations and the PE arrangement in perennial populations distributed over a wide swath of western North America. The AN arrangement of the inversion promoted rapid flowering over sustained vegetative growth and led to an annual life-history strategy that avoided summer drought. In contrast, the PE arrangement of the inversion promoted greater vegetative growth early in the season, followed by summer flowering and survival into subsequent years, and therefore a perennial life-history. The inversion was shown to have consistent effects on flowering time divergences in multiple independent population crosses through replicated quantitative trait locus (QTL) analysis. However, the inversion polymorphism's effect on flowering time was not great enough to overcome the large differences in flowering between the coastal perennial and inland annual genetic backgrounds across habitats. Hence the inversion did not contribute much to between habitat isolation caused by flowering time differences. Loci affecting traits other than flowering time are known to contribute to immigrant inviability between inland annual and coastal perennial populations. Nevertheless, the finding that inland perennial populations as well as coastal perennial populations had the PE arrangement of the inversion and that the early flowering selfing species *M. nasutus* had the AN arrangement, was further evidence that the distribution of the inversion was a function of the availability of soil moisture during summer months.

Genetic experiments (Lowry and Willis 2010) confirmed the extent of the inversion and the suppression of recombination (crossing over) in hybrids between 'AN' and 'PE'. No cross-over products were observed in more than 2270 effective meioses from inland annual  $\times$  coastal perennial crosses ( $F_2$ 's and RILs), so that the 95 % confidence interval for the recombination frequency of the inversion included a maximum 0.15 %. In contrast, recombination was found in crosses within the annuals and also in crosses within the perennials. The order of eight molecular markers in the annuals was inverted in the perennials. Markers in the inverted region spanned a genetic map region between the most distant markers e178 and e299 of at least 33 cM (33 % recombinants) in the mapping studies within the annual ecotype. While the inversion breakpoints are as yet unknown, the inversion encompasses at least 2.22 Mb and appears to contain 362 genes, identified from the draft genome sequence of *M. guttatus* ([www.phytozome.net](http://www.phytozome.net)) which comprises 450 Mb. These genes are inherited as a unit because of the lack of recombination.

The main conclusion for plant breeders is that they need to be aware of the possibility of inversions and their genetic consequences when working with cultivated and wild relatives of modern cultivars. As shown in Fig. 3.4, a crossover in a paracentric (not involving centromere) inversion results in a dicentric bridge and acentric fragment and almost certainly non viable recombinants. With a pericentric inversion, if the two breaks occur at different distances from the centromere, a shift

**Fig. 3.4** Paracentric inversion ( $C\dots J$  and  $j\dots c$ ) during meiosis. (a) Pairing of homologous chromosomes, heterozygous for paracentric inversion, results in loop. (b) Crossover in inversion results in dicentric bridge and acentric fragment. (c) Products of meiosis are, in theory, two non-recombinant chromatids (top and bottom), and two (non viable) chromatids with duplications and deletions



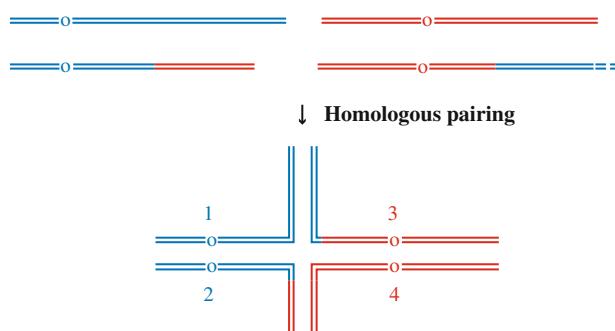
in the centromere position would occur. Crossing over within a pericentric inversion, when heterozygous, does not give rise to a dicentric bridge and acentric fragment, but the recombinant chromatids have duplications and deficiencies and again will almost certainly be non viable.

## Chromosomal Rearrangements in Wheat

Four hundred and sixty polyploid wheat accessions and 39 triticale forms from 37 countries of Europe, Asia, and USA were scored for chromosomal rearrangements by Badaeva et al. (2007) using C-banding staining of

heterochromatin in mitotic chromosomes. The germplasm included wild species, landraces, local varieties and commercial cultivars. The authors detected rearrangements in 70 of 208 accessions of tetraploid wheat, 69 of 252 accessions of hexaploid wheat, and 3 of 39 triticale forms. Altogether, 58 types of major chromosomal rearrangements were identified and discussed relative to 11 additional translocation types described by other authors. Six chromosome modifications of unknown origin were also observed. Single translocations were the most frequent type (39), followed by multiple rearrangements (9 types), pericentric inversions (9 types) and paracentric inversions (3 types). The breakpoints were located at or near the centromere in 60 rearranged chromosomes, while in 52 others they were in interstitial chromosome regions. In the latter, translocation breakpoints were often located at the border of C-bands and the euchromatin region or between two adjacent C-bands; some of these regions appeared to be translocation “hotspots”. The B-genome chromosomes were most frequently involved in translocations, followed by the A- and D-genome chromosomes; individual chromosomes also differed in the frequencies of translocations. Most translocations were detected in 1 or 2 accessions, and only 11 variants showed relatively high frequencies or were detected in wheat varieties of different origins or from different species. The authors concluded that high frequencies of some translocations with a very restricted distribution could be due to a “bottleneck effect”. Other types seemed to occur independently and their broad distribution could result from selective advantages of rearranged genotypes in diverse environmental conditions. The highest proportions of rearranged genotypes were found in Central Asia, the Middle East, Northern Africa and France. A low proportion of aberrant genotypes were characteristic of tetraploid wheat from Transcaucasia and hexaploid wheat from Middle Asia and Eastern Europe.

Four chromosomes, of which two non homologous ones have a reciprocal translocation (same colour represents homology in this figure)



Centromeres 1 and 4 and 2 and 3 need to go to same pole during anaphase I for balanced gametes.

**Fig. 3.5** Reciprocal translocation heterozygote: chromosome pairing during meiosis

Plant breeders need to be aware of the possibility of chromosomal rearrangements when utilizing landraces and wild relatives of cultivars in breeding and genetic studies. As a result of reciprocal translocations, genes could be linked in one parent but not the other, and chromosome pairing during meiosis could result in quadrivalents and irregular segregation. Changes in pairing partners are shown in Fig. 3.5. If chiasmata are formed in each of the paired arms, a ring of four chromosomes occurs at metaphase, whereas a chain of four results when one arm fails to form a chiasma. Their arrangement on the metaphase plate determines whether or not balanced gametes are formed without deficiencies and duplications. Balanced gametes are of two types: one having a ‘normal’ set of chromosomes, the other the translocation set.

## Chromosomal Evolution in the Plant Family Solanaceae

Most of this book is concerned with how variation within crop species and their wild relatives can be utilized by plant breeders. However, in this age of rapid and extensive next-generation sequencing of DNA, comparative genomics is being used to study plant evolution and also to aid the discovery of genes of economic importance in crop plants. It therefore seems appropriate to finish this chapter with an example, namely chromosomal evolution in the Solanaceae, which represents a group of dicotyledonous plants in the Euasterid clade, which is divergent from the model plant *Arabidopsis*. Wu and Tanksley (2010) combined data from multiple COSII (single-copy conserved orthologous markers) studies, and other comparative mapping studies performed in tomato, diploid potato, eggplant, pepper and diploid *Nicotiana* species (all  $2n = 2x = 24$ ), to deduce the features and outcomes of chromosomal evolution in the Solanaceae over the past 30 million years, based on the principle of parsimony. Orthologues are defined as genes sharing a common ancestor by speciation. Without going into details, Wu and Tanksley (2010) made use of a large set of orthologous sequences that had been aligned across species so that “universal PCR primers” could be designed, which amplified the corresponding orthologues from the different species. They constructed a molecular phylogenetic tree relative to the tomato-coffee split at 86 MYA (million years ago) in which tomato (T) and potato (Pt) diverged from their most recent common ancestor (ATPt) at 7.3 MYA; with the most recent common ancestors of eggplant (E) at 15.5 MYA (ATE), pepper (P) at 19.6 MYA (ATP) and *Nicotiana* (N) at 23.7 MYA (ATN). Comparing the maps of tomato, potato, eggplant and pepper genomes provided an opportunity to estimate rates of chromosomal evolution in the Solanaceae, resulting in an estimate of 0.1–1 inversions per million years and 0.2–0.4 translocations per million years across different species; in other words, 0.03–0.12 rearrangements per chromosome per million years. The authors concluded that the Solanaceae had experienced chromosomal changes at a modest rate compared with other families and that the rates were likely conserved across different lineages of the family. Chromosomal inversions occurred at a consistently

higher rate than translocations. Furthermore, they found evidence for non-random positioning of the rearrangement breakpoints suggesting that hot spots for chromosomal breakages played a significant role in shaping genome evolution. Finally, by utilizing multiple genome comparisons they were able to reconstruct the most likely genome configuration for a number of now-extinct progenitor species that gave rise to the extant solanaceous species studied in their research.

In summary, since these species diverged from their last most recent common ancestor, four inversions had occurred along the tomato lineage, two inversions along the potato lineage, four inversions along the ATPt lineage, 16 inversions along the eggplant lineage, and at least one inversion along each of ATE and pepper lineages as well as 11 undetermined inversions between these two lineages. The situation for translocations was more complicated. Since eggplant diverged from ATE, a T5 segment was inserted into E3, a T12 segment was inserted into E10, and markers from T10 were translocated to E4 and E10 respectively. Since pepper diverged from ATP, a non-reciprocal translocation resulted in P1 and P8, a small T4 segment was inserted into P12, and a small T12 segment was inserted into P3. In addition to these events with known timing, chromosomes 4, 5, 11 and 12 had been rearranged by translocations in the ATPt and/or eggplant genomes after they diverged from the ATE genome, and chromosomes 3, 4, 5, 9, 11 and 12 had been rearranged by translocations in the ATE and/or pepper genomes after they diverged from the ATP genome.

A more recent comparison of eggplant and tomato has revealed a slightly different picture for these species. Doğanlar et al. (2014) made a high resolution linkage map of 400 AFLP, 348 RFLP and 116 COSII markers (average spacing 1.8 cM) for 108 F<sub>2</sub> individuals from the interspecific cross of eggplant (*S. melongena*) MM738 and a wild relative *S. linnaeanum* MM195. The orthologous RFLP and COSII markers allowed confirmation of the established syntenic relationships between eggplant and tomato chromosomes. Thirty three rearrangements were identified between the eggplant and tomato linkage maps but only 14 were classified as inversions. The greater proportion was translocations: 13 simple translocations and 6 inverted translocations. Eight of the 19 translocations were interchromosomal. Only 2 of the 12 chromosomes, 6 and 8, showed no evidence of translocation events. In addition, 11 transpositions of single markers were detected with 6 of these being interchromosomal.

## Transposable Elements

Transposable elements (TEs) were discovered in maize in the 1940s by Barbara McClintock (1950). Kernels with different pigmentation from others in a cob indicated that a mutation(s) had changed pigment synthesis. More generally, a mobile DNA element can ‘jump’ into a structural gene and inactivate it. Subsequently it ‘jumps’ into another gene, whereby the function of the first gene is usually restored. These transposable elements were named transposons. It is

now known that the structure of the transposon Ac (activator) consists of double-stranded DNA of length 4600 bp (Heldt and Piechulla 2011). Both ends contain a 15 bp long inverted repeat sequence. Within the transposon, a structural gene encodes an enzyme transposase which catalyzes the transposition of the gene by binding to the two inverted repeats. In maize, another transposon (Ds) has been found in which the structural gene for the transposase is defective. The Ds transposon is therefore mobile only in the presence of the transposase from Ac. Another class of mobile elements is derived from retroviruses that have lost the ability to synthesize coat protein, and hence are called retrotransposons. They do not jump out of a gene but can multiply because at both ends of a retrotransposon there are signals for transcription by the host RNA polymerase. The resulting RNA encodes a reverse transcriptase that produces retrotransposon DNA (cDNA) which integrates into another section of the genome. Transposable (genetic) elements are now known to be ubiquitous in both prokaryotes and eukaryotes (Zhu et al. 2012). Genome sequencing has shown that TEs are the single largest component of the genetic material of most eukaryotes, and are considered by some to be a driving force of genome evolution. Retrotransposons are Class I elements which are further subdivided into long terminal repeat (LTR) retrotransposons, with both LTRs in direct repeat orientation, and non-LTR retrotransposons comprising long interspersed nuclear elements (LINEs) and short interspersed nuclear elements (SINES). DNA transposons are Class II elements, characterized by short terminal inverted repeats (TIRs) and duplication of a short genomic sequence at the site of insertion. Both groups of TEs contain autonomous and non-autonomous elements.

# Chapter 4

## Mendelian Genetics and Linkage Maps

### Introduction

The last chapter explained how the outcome of meioses determines the combinations of genes that are passed on to the next generation through sperm cells and egg cells and hence the results of hybridization experiments, whether for breeding or genetic studies. Plant breeders are therefore interested in the chromosomal locations of genes of economic importance in their crops. In this chapter we will explore the relationship between chromosomal locations and genetic linkage maps; but we need to start with Mendel's Laws of Inheritance and also take the opportunity to explore the idea of sampling variation arising by chance.

### Segregation

Mendel chose to investigate seven characters in the garden pea (*Pisum sativum*) which each displayed a clear cut (qualitative) difference between a pair of cultivars and for which the cultivars were true breeding, or in his words, "yielded perfectly constant and similar offspring". In each of his seven crosses, the hybrid resembled one parent, this parent having the dominant character and the other parent the recessive character, irrespective of whether the parent was used as the seed or pollen plant. In the next generation, produced by self-pollination of the hybrid, he found on average a three to one ratio of offspring with the dominant and recessive character, respectively. All of the offspring with the recessive character bred true whereas on average one third of those with the dominant character did likewise, but two thirds behaved like the hybrid and displayed the same three to one ratio in their progeny. Mendel symbolized his law of segregation as follows. If  $A$  be taken as the constant dominant character (e.g. round peas),  $a$  the constant recessive character (wrinkled peas), and  $Aa$  the hybrid form (round peas) in which both are conjoined,

the ratios of the types of progeny from a hybrid between the two forms of the differentiating character are as follows.

$$A + 2Aa + a$$

After  $n$  generations of self-pollination the ratios of  $A$  to  $Aa$  to  $a$ , would be  $2^n - 1$  to  $2$  to  $2^n - 1$ . Today we call the hybrid the  $F_1$  generation and hence  $n$  generations of self-pollination results in the  $F_{n+1}$  generation.

## Independent Assortment

Mendel then made two experiments with a considerable number of plants in which he examined the offspring of hybrids in which two and three differentiating characters, respectively, were associated. If, for simplicity, the different characters of the seed parent are indicated by  $A$ ,  $B$  and  $C$ ; then those of the pollen parent are indicated by  $a$ ,  $b$  and  $c$ ; and the hybrid forms for the characters are  $Aa$ ,  $Bb$  and  $Cc$ . Mendel found that the full number and ratios of the classes of offspring from hybrids with two kinds of differentiating characters were given by the terms in the combinations of the expressions for each kind:

$$(A + 2Aa + a)(B + 2Bb + b)$$

And likewise for three kinds of differentiating characters:

$$(A + 2Aa + a)(B + 2Bb + b)(C + 2Cc + c)$$

Thus the segregation of one character difference is independent of any other character difference; Mendel's law of independent assortment. If  $n$  represents the number of the differentiating characters in the two original stocks,  $3^n$  gives the number of terms of the combination series,  $4^n$  the number of individuals which belong to the series and  $2^n$  the number of unions which remain constant (i.e. true breeding). Mendel obtained all 128 ( $2^7$ ) constant combinations that were possible from his seven differentiating characters. Mendel's explanation of his results was: "that in the ovaries of the hybrids there are formed as many sorts of egg cells, and in the anthers as many sorts of pollen cells, as there are possible constant combination forms, and that these egg and pollen cells agree in their internal composition with those of the separate forms". Mendel added "that the various kinds of egg and pollen cells were formed in the hybrids on average in equal numbers". Thus with one differentiating character, each pollen form  $A$  and  $a$  unites equally often with each egg cell form  $A$  and  $a$  to give:

$$A/A + A/a + a/A + a/a = A + 2Aa + a.$$

Mendel confirmed his explanation through the ratios found in backcross generations. Thus when the round yellow hybrid between round yellow ( $AB$ ) and wrinkled green ( $ab$ ) peas was crossed with pollen of true-breeding round yellow peas, all of the offspring were round yellow, whereas when they were crossed with pollen of true-breeding wrinkled green peas, the result was nearly equal numbers of round yellow, round green, wrinkled yellow and wrinkled green seeds. Departures from exactly equal numbers were attributed to chance fluctuations in small samples of offspring. The same was true in the reciprocal crosses. Thus in the hybrids, egg and pollen cells of the forms  $AB$ ,  $Ab$ ,  $aB$  and  $ab$  were produced in equal numbers. In other words, Mendel's pea hybrids formed "egg and pollen cells which, in their constitution, represent in equal numbers all constant forms which result from the combination of the characters united in fertilization".

It was Walter Sutton who in 1903 (Sutton 1903) correctly explained Mendel's laws in terms of chromosome behaviour (pairing and separation) during the two nuclear (cell) divisions of meiosis. Thus random orientation of different chromosome-pairs on the first-division spindle results in independent assortment, whilst segregation occurs either with the separation of centromeres and chromosomes to opposite poles of the spindle during anaphase I or with the separation of chromatids to opposite poles during anaphase II. In the chapter on gene expression we shall return to some of the genes studied by Mendel.

## The Binomial Distribution

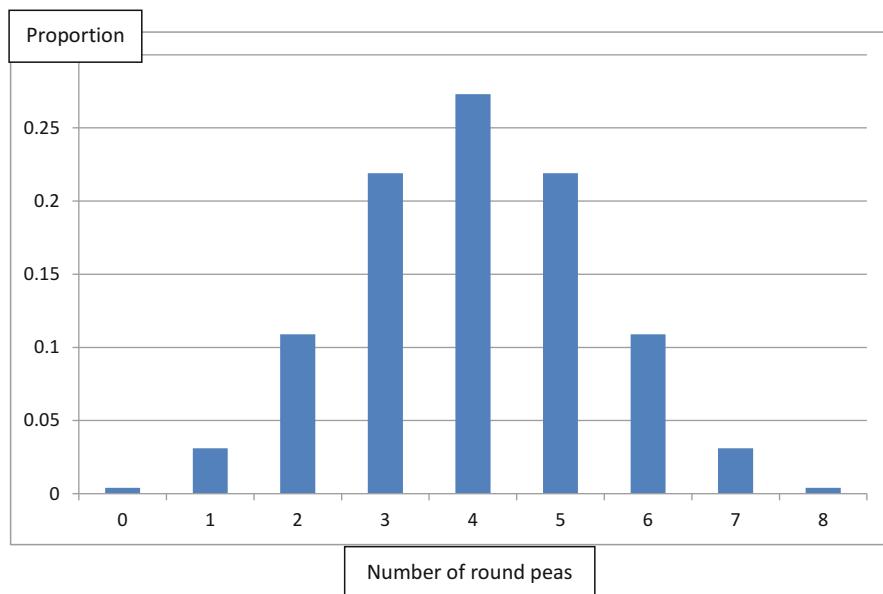
Mendel attributed departures from his expectations to chance fluctuations in small samples of offspring. In the backcross mentioned above, he expected a 1:1 ratio of round to wrinkled peas. This is like tossing a coin and expecting a 1:1 ratio of heads to tails. In a sample of size  $n$ , with the probability of a round pea ( $P$ ) equal to the probability of a wrinkled one ( $Q$ ), and hence both equal to  $\frac{1}{2}$ , one can use the binomial distribution to work out the probabilities of all of the possible ratios.

Let's start with a small sample size of just eight observations. The probability of all wrinkled peas is  $\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2}$ . The probability of one round pea and seven wrinkled peas is the same, but there are eight ways of obtaining one round pea, namely the first member of the sample, or the second member, or indeed any member up to number eight. The complete set of probabilities is as shown in Table 4.1.

Hence the probability of exactly four round and four wrinkled peas in a sample is 0.2734375. This is also the proportion of samples, in a very large number, expected to have this ratio of round to wrinkled peas, and the same is true for the other ratios and proportions shown in Table 4.1 and displayed in Fig. 4.1.

**Table 4.1** Probabilities of 0–8 round peas in a sample of 8

Round	Wrinkled	Probability	$P = Q = \frac{1}{2}$
0	8	$Q^8$	0.00390625
1	7	$8PQ^7$	0.03125
2	6	$28P^2Q^6$	0.109375
3	5	$56P^3Q^5$	0.21875
4	4	$70P^4Q^4$	0.2734375
5	3	$56P^5Q^3$	0.21875
6	2	$28P^6Q^2$	0.109375
7	1	$8P^7Q$	0.03125
8	0	$P^8$	0.00390625



**Fig. 4.1** Graphical display of proportion of samples of size 8 with a given number of round peas

The mean of this distribution (the mean number of round peas in a sample of 8) can be calculated as follows.

$$0 \times 0.00390625 + 1 \times 0.03125 + \dots + 8 \times 0.00390625 = 4$$

The mean is 4 as might have been expected. The variation from sample to sample (from 0 up to 8 round peas in a sample of 8) can be summarized as the variance of the distribution which is calculated as follows.

$$0.00390625 \times (0 - 4)^2 + 0.03125 \times (1 - 4)^2 + \dots + 0.00390625 \times (8 - 4)^2 = 2$$

**Table 4.2** Comparison of binomial and chi-square probabilities for sample size of 8

Binomial		Chi-square			
Ratio	Probability	Ratio	$\chi^2$	Probability	
4:4	or worse	1.0000	4:4	0.00	1.00
3:5 or 5:3	or worse	0.7265625	3:5 or 5:3	0.50	0.50–0.30
2:6 or 6:2	or worse	0.2890625	2:6 or 6:2	2.00	0.20–0.10
1:7 or 7:1	or worse	0.0703125	1:7 or 7:1	4.50	0.05–0.02
0:8 or 8:0		0.0078125	0:8 or 8:0	8.00	0.01–0.001

More generally, for a sample of size  $n$ , the probability of  $x$  round peas ( $P(x)$ ) is

$$n! / [(x!(n-x)!)] \times P^x Q^{(n-x)}$$

The average number (or expected value) of round peas is  $nP$ , and the variance of the number of round peas is  $nPQ$ . The variance increases as  $n$  increases. However, if the proportion of round peas is considered (0, 1/ $n$ , 2/ $n$  ... 1), with the appropriate binomial probabilities, the mean is  $P$  (=½) and the variance is  $PQ/n$  (=1/(4n)).

Now returning to our specific example, we need to ask the question, how big a departure from 4 round and 4 wrinkled peas can be accepted as due to chance. We have seen that the probability of exactly four round and four wrinkled peas in a sample is only 0.2734375, but an experimenter would be very happy to obtain four observations in each class! Hence a better way of comparing observed with expected ratios is the one shown in Table 4.2.

It can be seen that with such a small sample, a ratio of 1 to 7 or 7 to 1 or worse (i.e. 0 to 8 or 8 to 0) can occur seven times in every 100 experiments; not that rare an event. Mendel, however, had a sample size of 110 in one of his backcrosses, and observed 57 round and 53 wrinkled peas. With larger sample sizes it is tedious for us (although not for a computer) to work out the exact probabilities, but fortunately there is goodness-of-fit statistic (**chi-square** ( $\chi^2$ )) which is a very good approximation, particularly as the sample size increases.

## Chi-Square ( $\chi^2$ ) Goodness-of-Fit Statistic

The **chi-square** ( $\chi^2$ ) goodness-of-fit statistic is in fact a commonly used test of significance devised by Karl Pearson in 1900 (Bulmer 1967) for categorical data. If a category has an observed count of  $o$  observations and an expected count  $e$  under some hypothesis, then the goodness-of-fit test statistic is formed by summing the quantity  $(o - e)^2/e$  over categories:

$$X^2 = \sum (o - e)^2/e, \text{ where summation is over categories.}$$

If the hypothesis is true, it is likely that  $o$  will differ from  $e$  by chance in each category, but not by too much, as can be seen for our simple example in Table 4.2. How far apart they may be by chance can be found from the properties of the chi-square distribution. In particular, for two categories there is a 5 % chance that  $X^2$  will exceed 3.84 if the hypothesis is true, a 1 % chance that it will exceed 6.63 and a 0.1 % chance that it will exceed 10.83. And that does mean that if you do 1000 tests, you can expect one to exceed 10.83 by chance. These chi-square values can be looked up in tables such as those compiled by Fisher and Yates (1963), remembering that for two categories there is just one degree of freedom (1df) because for a given number of observations, once you know the numbers in one category you know the numbers in the second category. The probabilities for our simple example are given in Table 4.2. An 8 to 0 or 0 to 8 ratio with a chi-square of 8.0 would be rejected as unlikely by chance (probability of 0.01–0.001, or 0.0078125 exactly), whereas 1 to 7 or 7 to 1 or worse with a chi-square of 4.50 is borderline (probability 0.05–0.02, but a slightly different exact value of 0.070).

We can now return to Mendel's backcross. Mendel found 31 round and yellow, 26 round and green, 27 wrinkled and yellow, and 26 wrinkled and green seeds. The chi-square test for a 1 to 1 ratio of round to wrinkled seeds is:

$$(57 - 55)^2/55 + (53 - 55)^2/55 = 8/55 = 0.145 \text{ which is much less than 3.84 and hence a good fit.}$$

The chi-square test for a 1 to 1 ratio of yellow to green seeds is:

$$(52 - 55)^2/55 + (58 - 55)^2/55 = 18/55 = 0.33 \text{ which is also much less than 3.84 and hence a good fit.}$$

The chi-square test, with three degrees of freedom, for a 1:1:1:1 ratio for the four classes is:

$$(31 - 27\frac{1}{2})^2/27\frac{1}{2} + (26 - 27\frac{1}{2})^2/27\frac{1}{2} + (27 - 27\frac{1}{2})^2/27\frac{1}{2} + (26 - 27\frac{1}{2})^2/27\frac{1}{2} = 17/27\frac{1}{2} = 0.62 \text{ which is much less than 7.81 (5 % for 3df) and hence a very good fit.}$$

Subtracting the two chi-squares for the 1:1 segregations, each with one degree of freedom, from the chi-square with three degrees of freedom, gives a chi-square of  $(0.62 - 0.145 - 0.33) = 0.145$  with one degree of freedom. This is the test for independent assortment, and with a value much less than 3.84 is a good fit.

In the above example there was no evidence of departures from normal segregation at the two loci, but if there was evidence of distorted segregation, one can still use a chi-square test of independent assortment through a contingency table where the segregation ratio at each locus is estimated, leaving 1 df for testing for independence (Box 4.1). Again, using the numbers from Mendel's backcross we obtain a  $X^2_{AB}$  of 0.13, which is much less than 3.84 and hence a good fit to expectation with independent assortment.

**Box 4.1: Contingency Table ( $2 \times 2$ )**

Observed

Locus	<i>B</i>	<i>b</i>	Totals
<i>A</i>	<i>a</i>	<i>b</i>	<i>a</i> + <i>b</i>
<i>a</i>	<i>c</i>	<i>d</i>	<i>c</i> + <i>d</i>
Totals	<i>a</i> + <i>c</i>	<i>b</i> + <i>d</i>	<i>n</i>

Expected

Locus	<i>B</i>	<i>b</i>	Totals
<i>A</i>	$(a+b)(a+c)/n$	$(a+b)(b+d)/n$	<i>a</i> + <i>b</i>
<i>a</i>	$(c+d)(a+c)/n$	$(c+d)(b+d)/n$	<i>c</i> + <i>d</i>
Totals	<i>a</i> + <i>c</i>	<i>b</i> + <i>d</i>	<i>n</i>

The contribution of the ‘*a* cell’ to the  $\chi^2$  statistic is  $[a - (a+b)(a+c)/n]^2 / [(a+b)(a+c)/n] = (ad - bc)^2 / [n(a+b)(a+c)]$ . The other cells make similar contributions and summation gives the following test for independence with one degree of freedom.

$$\chi^2_{AB} = n(ad - bc)^2 / [(a+b)(c+d)(a+c)(b+d)]$$

## Terminology

Following the rediscovery of Mendel’s work in 1900, the scientific study of heredity and variation made rapid progress in many organisms, and was named ‘Genetics’ by the British biologist William Bateson in his presidential address to the Third International Conference organized by the Royal Horticultural Society (RHS) in 1906, which was renamed the Third International Conference on Genetics. He referred to the contrasting forms *A* and *a* as allelomorphs, later trimmed to alleles. He introduced the  $P_1$ ,  $F_1$  and  $F_2$  symbols for the parental and filial (offspring) generations. He replaced Mendel’s symbolism  $A:2Aa:a$  with  $AA:2Aa:aa$  and referred to the genetic constitution as homozygous ( $AA$  or  $aa$ ) or heterozygous ( $Aa$ ). Later, in 1909, the Danish biologist Wilhelm Johannsen named the heredity units, genes, and introduced the terms genotype (genetic makeup, such as  $AA$ ) and phenotype (the associated characteristic, such as round peas). It quickly became apparent that the number of alternatives for a particular character which showed Mendelian inheritance was not limited to two. In other words, multiple alleles occur with one of the first examples in plants being cob and pericarp colour in maize (Emerson 1911).

As more traits were studied in a wide range of organisms it became clear that independent assortment did not apply to all pairs of character differences and that a particular phenotype could be the result of interactions between the alleles of different genes.

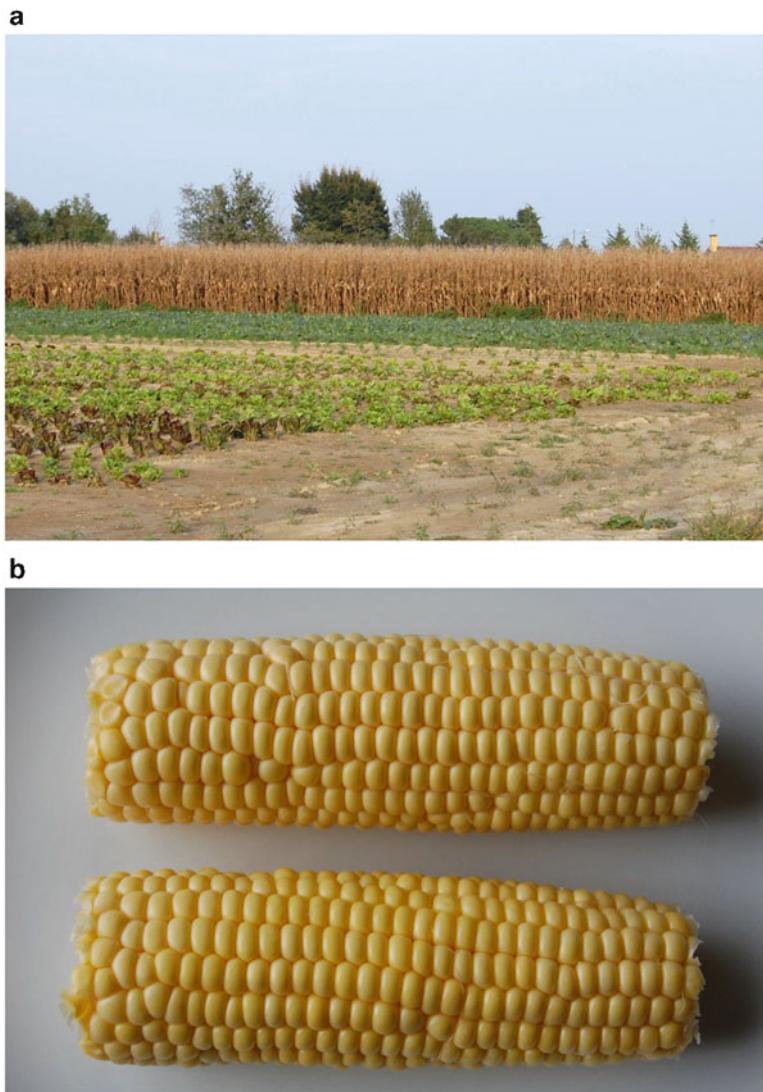
## Partial Linkage and Linkage Maps

Bateson et al. (1905), working with sweet peas (*Lathyrus odoratus*), were the first to discover an exception to the law of independent assortment with the character-differences purple (*P*) versus red (*p*) flowers and long (*L*) versus round (*l*) pollen. The F<sub>2</sub> combinations from the cross Purple Long × red round were 284 (214) *PL*, 21 (71½) *Pl*, 21 (71½) *pL* and 55 (24) *pl*; with *PL* and *pl* in excess of expectation (in parentheses) from independent assortment. The contingency chi-square (1df) for departure from independent assortment (i.e. linkage) is 163.37, considerably larger than the 10.83 required for significance at the 0.1 % level. This coupling of ‘Purple’ with ‘Long’ implied that the frequencies of *PL* and *pl* were appreciably higher than *Pl* and *pL* at the time of pollen and egg-cell formation in the F<sub>1</sub> plants; in other words, they were partially linked. Bateson used the term repulsion when Purple and Long were in different parents; Purple round × red Long.

The chromosomal explanation of partial linkage, and the extension of the work to the construction of genetic linkage maps, came from Thomas Hunt Morgan and his fruitfly (*Drosophila melanogaster*) geneticists at Columbia University in the USA. They concluded that certain genes tend to remain together because they lie near each other in the same chromosome; the closer together the greater the association. They introduced the term *crossing-over* for the process of interchange by which new combinations of linked factors arise (Morgan and Cattell 1912). It was Morgan’s student Alfred Sturtevant who at the age of 21 produced the first genetic map (linear arrangement of genes in linkage group) of six sex-linked genes in the fruitfly (Sturtevant 1913). The unit of distance between two genes was named after Morgan, and is defined such that in one Morgan (M), on average, one cross-over will occur in every gamete formed so that in one centi-Morgan (cM), on average, one cross-over will occur for every 100 gametes formed. Morgan favoured Janssens’ theory as the explanation of genetic cross-overs. Since each bivalent (tetrad) consists of four chromatids, and each cross-over involves two, the (genetic) map distance between two loci is half the average number of chiasmata so that one chiasma equals 0.5 M. Hence a chromosome-pair with an average of two chiasmata will have a length of 1 M or 100 cM.

## Maize

The plant that proved particularly useful for linkage studies was also an economically important one, maize (*Zea mays*) (Fig. 4.2). Bregger (1918) crossed a pure-breeding strain with coloured (dominant) aleurone and starchy (dominant) endosperm (*CC WxWx*) with one having colourless aleurone and waxy endosperm (*cc wxwx*) so that the doubly heterozygous F<sub>1</sub> had the genes in coupling (*C Wx/c wx*). When this was test-crossed (backcrossed) with *cc wxwx*, the resulting ear had 403 grains of which 147 were coloured and starchy (*C Wx*), 65 coloured and



**Fig. 4.2** Maize, economically valuable and useful for genetic research. (a) Maize ready for harvest, near Venice, Italy (5 October 2014). (b) Sweet corn cooked and ready to eat

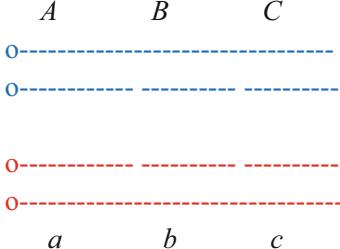
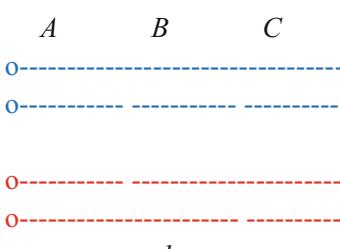
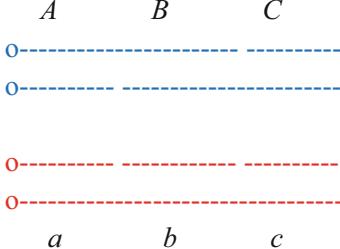
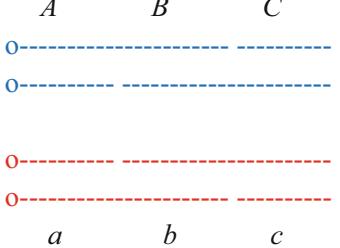
waxy ( $C\ wx$ ), 58 colourless and starchy ( $c\ Wx$ ) and 133 colourless and waxy ( $c\ wx$ ). Hence there was no evidence of a departure from 1 to 1 segregation for  $C$  and  $c$  ( $X^2_C = 1.09; P > 0.05$ ), nor for  $Wx$  and  $wx$  ( $X^2_{Wx} = 0.12; P > 0.05$ ), but the chi-square test for linkage was highly significant ( $X^2_{CWx} = 61.16; P < 0.001$ ); and so was the contingency chi-square which had a similar value (61.06), as expected

given that there was normal segregation at each locus. The observed cross-over (recombination) frequency was 30.5 % [(65 + 58)/403] so that the estimated distance between loci was 30.5 cM. Bregger (1918) obtained a slightly lower estimate (26.7 cM) of recombination frequency when the genes were in repulsion ( $C\ wx/cWx$ ), but the two estimates were not significantly different, as can be shown with a contingency chi-square test. Shortly we will consider the question of how good these estimates are of recombination frequency.

Work on maize also provided one of the proofs that the cytological (physical) crossing-over of homologous chromosomes is accompanied by genetic crossing-over of genes in the same linkage group. Creighton and McClintock (1931) bred a strain of maize that was heterozygous for two structural characteristics of chromosome 9 (presence versus absence of knob and presence versus absence of a segment of chromosome 8) and for the two endosperm characters (coloured versus colourless aleurone layer and starchy versus waxy endosperm). Cytological and genetic crossing-over occurred together when this strain was used in an appropriate cross. Darlington (1934) showed that there was agreement between chiasma and cross-over frequencies for each chromosome of maize, in which there was a mean total chiasma frequency of 27.1 for the ten chromosome-pairs in the pollen-mother-cells. Subsequently Rhoades (1950) provided more extensive data on meiosis in maize. By the 1950s textbooks of genetics (e.g. Sinnott et al. 1958) were showing the locations of several hundred gene loci on the 10 linkage groups of maize corresponding to its 10 microscopically visible chromosomes. Before dealing with the construction of linkage maps, it is useful to consider the results of ordered tetrad analysis.

## Ordered Tetrad Analysis

As mentioned in Chap. 3, it was ascomycete fungi such as *Neurospora crassa* and *Sordaria fimicola* (Lindgren 1933; Whitehouse 1973) that proved of tremendous value for studying meiosis because ordered tetrad analysis was possible. Ascomycete fungi are haploid and exist in two mating-types. Meiosis occurs immediately after fusion of the nuclei of opposite mating-type. The four resulting haploid nuclei are linearly arranged within an elongated tube-like cell, the ascus, such that the two second-division spindles do not overlap. Subsequently mitosis takes place in each nucleus, again without spindle overlap, and spore walls form around each nucleus. The result is a mature ascus with a linear sequence of eight elliptical spores, in which the members of each pair are genetically identical, and the four pairs are the four meiotic products (Whitehouse 1973). With three or more linked genes the four tetrad genotypes are all different. It was therefore possible to obtain direct evidence that crossing-over occurs at the four-strand stage and involves only two of the four strands at any one place (Fig. 4.3), and furthermore that the paternal and maternal centromeres of the chromosome pair segregated from one another at the first division of meiosis. It was also possible to explore double cross-overs and to

(i) Two-strand	Tetrad	Number of crossovers
	o <i>A B C</i>	0
	o <i>A b C</i>	2
	o <i>a B c</i>	2
	o <i>a b c</i>	0
(ii) Three-strand	Tetrad	Number of crossovers
	o <i>A B C</i>	0
	o <i>A b c</i>	1
	o <i>a B c</i>	2
	o <i>a b C</i>	1
(iii) Three-strand	Tetrad	Number of crossovers
	o <i>A B c</i>	1
	o <i>A b C</i>	2
	o <i>a B C</i>	1
	o <i>a b c</i>	0
(iv) Four-strand	Tetrad	Number of crossovers
	o <i>A B c</i>	1
	o <i>A b C</i>	1
	o <i>a B C</i>	1
	o <i>a b C</i>	1

**Fig. 4.3** The four kinds of double cross-overs from hybrid  $ABC/abc$  (haploid  $ABC \times abc$  or diploid  $ABC/abc$ , whose gametes would be combined with  $abc$  gametes from  $abc/abc$  through a three-point testcross). Blue and red: two homologous chromosomes with two chromatids (circles are centromeres). Breaks are cross-overs.

investigate cross-over position interference and chromatid interference. If the four kinds of double cross-over are not equally frequent, one has chromatid interference, or in other words, non-random distribution occurs between the chromatids involved in successive cross-overs. Whitehouse (1973) concluded that there is usually only slight chromatid interference so that any genetic interference between cross-overs must often be attributed to the position of one point of exchange interfering with the position of a neighbouring exchange, i.e. cross-over position interference. An extension of Fig. 4.3 to include many loci would also demonstrate that the maximum amount of recombination between two loci is 50 %. Between loci A and B there are equal numbers of chromatids with zero (no recombination) and one (recombination) cross-over; between A and C there are equal numbers of chromatids with even (zero and two: no recombination) and odd (one: recombination) numbers of cross-overs; and the same is true as more loci are added. In other words, with one or more chiasmata between loci, the recombination frequency is 50 %. Hence in contrast to genetic map distance, recombination frequencies are not additive (i.e. A-C is not the sum of A-B and B-C).

## ***Interference***

Returning to maize, Stadler (1926) located a third endosperm gene (*Sh/sh*), for convex (dominant) versus shrunken concave surface, between *C/c* and *Wx/wx* and was able to provide an example of double crossing-over between these two genes. The number of double-crossovers observed (32) was less than that expected (212.4) if the two cross-overs were occurring at random. The ratio (32/212.4 = 0.151) is called the *coincidence* (Muller 1916).

Given the existence of cross-over position interference, it is reasonable to assume that the underlying cause is chiasma position interference. Haldane (1931) demonstrated this from data on chiasma frequencies at metaphase I in the six chromosome-pairs of *Vicia faba*. He pointed out that if the distribution of chiasmata was at random along the chromosomes, the frequencies of chromosomes with different numbers of chiasmata should fit a Poisson distribution in which the variance was equal to the mean. However, for the five short chromosome-pairs the variance was only about one quarter of the mean (mean = 3.5) and for the long chromosome-pair it was about one third (mean = 8.1). The occurrence of one chiasma was greatly reducing the likelihood of another. Previously, Haldane (1919) had proposed a mapping function based on random distribution to convert cross-over (recombination) frequencies into additive map distances. Subsequently, mapping functions were proposed to take account of interference, of which Kosambi's (1944) is most widely used.

It is now time to look at the estimation of recombination frequencies, mapping functions, and the construction of linkage maps. The opportunity is taken to introduce the concept of maximum likelihood estimation.

## Estimation of Recombination Frequency

### Backcross

The observed two-locus counts for a backcross ( $AB/ab \times ab/ab$ ) and the expected counts, for normal segregation at each locus and a recombination frequency  $r$  between loci, can be written as follows:

		Offspring				Total
Genotype	$AB/ab$	$Ab/ab$	$aB/ab$	$ab/ab$		
Phenotype	AB	Ab	aB	ab		
Observed	$a$	$b$	$c$	$d$	$n$	
Expected	$n(1-r)/2$	$nr/2$	$nr/2$	$n(1-r)/2$	$n$	
Probability	$(1-r)/2$	$r/2$	$r/2$	$(1-r)/2$	1	

In a very large sample of gametes the proportion of each type of recombinant offspring is  $r/2$  by definition, making a total proportion of  $r$  recombinants, and  $1 - r$  non recombinants. Hence the probability of a recombinant is  $r$ . If one takes lots of samples of gametes of size  $n$ , one gets a binomial distribution in which the average number of recombinants is  $nr$  and the variance of the number of recombinants is  $nr(1 - r)$ . If the proportion of recombinants is considered, then the mean of the distribution is  $r$  and its variance is  $r(1 - r)/n$ .

In a particular sample of size  $n$ , an obvious estimate of  $r$  is  $(b + c)/n$  with variance  $r(1 - r)/n$ . It can be shown that this estimate of  $r$  is the maximum likelihood estimate (MLE). The method of maximum likelihood is a very general method of estimation which was developed by Ronald A. Fisher and leads to estimators with a number of desirable properties (Bulmer 1967). The method of maximum likelihood consists of taking as estimators, those values of the parameters which maximize the likelihood of the observations. In practice, for ease of computation, the logarithm of the likelihood ( $l$ ) is maximized; its maximum occurs at the same value of a parameter as that of the likelihood ( $L$ ). The details need not concern us, but it is useful to consider the likelihood (probability) of  $(b + c)$  recombinants in a sample of size  $n$ , given the unknown parameter  $r$ .

$$\text{The likelihood } L \text{ is} [n!/\{(b + c)!(a + d)!\}] \times r^{(b+c)}(1 - r)^{(a+d)}$$

Hence the ratio ( $LR$ , called the odds) of the likelihood (probability) that  $r$  equals the estimate of  $r$  ( $\hat{r}$ ) over the likelihood (probability) that  $r$  equals  $\frac{1}{2}$  (its maximum value) is  $L(r = \hat{r})/L(r = \frac{1}{2}) = \hat{r}^{(b+c)}(1 - \hat{r})^{(a+d)}/(\frac{1}{2})^n$ , as  $a, b, c$  and  $d$  are the same observed numbers in each situation so that  $[n!/\{(b + c)!(a + d)!\}]$  cancels out.

The significance of the odds ratio can be tested through its deviance ( $D$ ), which is twice the natural logarithm of the likelihood ratio:  $D = 2\ln(LR)$ . Under the null hypothesis of no linkage and for large values of  $n$ , the deviance follows a  $\chi^2$  distribution with one degree of freedom. However, the LOD score is more commonly used.

The common logarithm to base 10 of the likelihood ratio (the LOD) is:

$$(b + c)\log\hat{r} + (a + d)\log(1 - \hat{r}) - n\log\left(\frac{1}{2}\right)$$

If the odds are greater than 1000 (LOD = 3) one has very good evidence for linkage, but when considering a large number of pairs of loci, a more stringent criterion may be required.

In the maize example given earlier, the recombination frequency between *C/c* and *Wx/wx* was 0.305 cM. We can now calculate the standard error (square root of the variance) of this estimate as  $[0.3(1 - 0.3)/403]^{1/2} = 0.023$ , and the LOD score as  $123\log 0.3 + 280\log 0.7 - 403\log 0.5 = -64.314 - 43.373 + 121.315 = 13.628$ , which is considerably larger than 3, and hence extremely strong evidence for linkage.

Using likelihood theory, the approximate variance (var) of a parameter can be obtained as the inverse of the (Fisher) information, the latter being the expected value (*E*) of minus the second derivative of the log-likelihood (*l*). The expected value of the number of recombinants is *nr* and of non-recombinants *n(1 - r)*. For  $\hat{r}$  it can be shown that:

$$\text{var}(\hat{r}) = 1/E[-d^2l/dr^2] = \hat{r}(1 - \hat{r})/n$$

This result was derived above from the properties of the binomial distribution. If the estimate of *r* is zero, its true value may be more than zero so that the estimated variance of zero gives a false sense of accuracy.

### ***Complications and the EM Algorithm***

The backcross analysis was straightforward because the counts of phenotypic classes were also counts of genotypic classes and the types of gametes. In other words, we were able to count the number of recombinant gametes and divide this by the total number of gametes to estimate the recombination frequency. Analysis of other types of mapping population can be more complicated, with a segregating F<sub>2</sub> population providing a good example.

Let us consider *AB/AB* × *ab/ab* (grandparents) giving an *AB/ab* F<sub>1</sub> (parents) which is selfed or intercrossed to give a segregating F<sub>2</sub> offspring population. The following set of F<sub>1</sub> male gametes (probabilities) *AB* (1 - *r*)/2, *Ab* *r*/2, *aB* *r*/2 and *ab* (1 - *r*)/2 is combined at random with the same set of F<sub>1</sub> female gametes *AB* (1 - *r*)/2, *Ab* *r*/2, *aB* *r*/2 and *ab* (1 - *r*)/2 to give 16 offspring genotypes, but only ten are distinct and furthermore only nine phenotypes can be recognized (codominant alleles at both loci). Hence we encounter a difficulty when we try to count the number of recombinant gametes, as shown in Table 4.3. The problem is that we can count the number of offspring with AaBb phenotype but they comprise two genotypes, *AB/ab* (*A* and *B* alleles in coupling) and *Ab/aB* (*A* and *B* alleles in

**Table 4.3** The number of recombinant gametes in F<sub>2</sub> offspring (phenotype AABB is formed from an AB gamete uniting with an AB gamete to give unique AABB genotype, whereas AaBb phenotype can be formed from non-recombinant gametes AB + ab and recombinant ones Ab + aB, with probabilities of  $\frac{1}{2}(1 - r)^2$  for AB/ab and  $\frac{1}{2}r^2$  for Ab/aB genotypes, respectively, so that this is the ratio of non-recombinant to recombinant gametes, namely  $p_{co}$  to  $p_{re}$ )

Phenotypes	Observed number and example	Non-recombinant gametes	Recombinant gametes	Number recombinant gametes
AABB	$n_1$ 40.5	$2n_1AB$	—	0
AABb	$n_2$ 9.0	$n_2AB$	$n_2Ab$	9
AAAb	$n_3$ 0.5	—	$2n_3Ab$	1
AaBB	$n_4$ 9.0	$n_4AB$	$n_4aB$	9
AaBb	$n_5$ 82.0	$2n_5p_{co}$	$2n_5p_{re}$	$164p_{re}$
Aabb	$n_6$ 9.0	$n_6ab$	$n_6Ab$	9
aaBB	$n_7$ 0.5	—	$2n_7aB$	1
aaBb	$n_8$ 9.0	$n_8ab$	$n_8aB$	9
aabb	$n_9$ 40.5	$2n_9ab$	—	0
Total	$n$ 200.0			$(38 + 164p_{re})/400$

repulsion), and are therefore uninformative about  $r$ . The number of non-recombinant and recombinant gametes they contribute is  $2n_5p_{co}$  and  $2n_5p_{re}$ , respectively, where  $p_{re} = r^2/[(1 - r)^2 + r^2]$  and  $p_{co} = 1 - p_{re}$ . Hence the recombination frequency  $r$  (number of recombinant gametes divided by total of  $2n$ ) is:

$$r = (n_2 + 2n_3 + n_4 + 2n_5p_{re} + n_6 + 2n_7 + n_8)/(2n) \text{ where } p_{re} \text{ is a function of } r.$$

If we can't estimate  $r$  by simply counting the number of recombinant gametes, perhaps we can use the method of maximum likelihood. Unfortunately, the result is a polynomial equation without an explicit solution. Nevertheless, a maximum likelihood solution can be obtained using the 'expectation maximization' (EM) algorithm (Dempster et al. 1977). The algorithm is an iterative process in two steps. First the expectation is determined regarding the incomplete information and then the maximum likelihood estimate is computed under the assumption that the incomplete information is fixed (constant) as expected in the first step. A new expectation step is then taken and the process continued. Under most circumstances the algorithm will converge towards the maximum likelihood estimates of the unknown parameters. Let us look at a simple example with our F<sub>2</sub> population. Some actual numbers are shown in Table 4.3 so that the number of recombinant gametes divided by the total is  $(38 + 164p_{re})/400$ . As we don't know the value of  $r$ , let's start with a value of 0.25, halfway between 0 and 0.5, its lowest and highest possible values. We can now work out the expected value of  $p_{re}$  and hence the expected number of recombinant gametes. This allows us to obtain a new maximum likelihood value of  $r$ , namely the expected number of recombinant gametes divided by the total number of gametes. This new value of  $r$  leads to a new expected value of  $p_{re}$  and hence a new expected number of recombinant gametes, and so on. It can be seen in Table 4.4 that  $r$  rapidly converges to 0.1, the maximum likelihood estimate of  $r$ . In practice, fewer than ten iterations are usually required. A good account and more details of the

**Table 4.4** EM algorithm applied to example with  $F_2$  population of 400 gametes and observed numbers of phenotypes in Table 4.3

Iteration	$r$	Number recombinant gametes	$\hat{r}$
1	0.25	$38 + 16.4 = 54.4$	0.136
2	0.136	$38 + 3.965 = 41.965$	0.105
3	0.105	$38 + 2.2266 = 40.2266$	0.1006
4	0.1006	$38 + 2.0264 = 40.0264$	0.100066
Convergence	0.1	$38 + 2 = 40$	0.1

theory and practice of genetic mapping can be found in the book *Genetic Mapping in Experimental Populations* by Van Ooijen and Jansen (2013). These authors go on to consider the three situations involving dominance in an  $F_2$ , namely dominance at one locus and dominance at both loci with the dominant alleles in coupling and in repulsion. The accuracy of the estimates of  $r$  (information available about  $r$  in different situations) leads them to the conclusion that reliable map orders of dominant markers in repulsion phase with each other are virtually impossible to achieve in practice. They also consider populations that have passed through multiple rounds of meiosis with segregation and recombination (e.g. recombinant inbred lines derived from an  $F_1$  by single seed descent) and outbreeding species. The genetics of outbreeding species are more complicated than those of inbreeding species. Outbreeding means that two genetically distinct parents contribute the male and female gametes that combine to form the offspring of the mapping population. Up to four different alleles may be present at each locus and the linkage phases may vary across loci and between the parents. Furthermore, recombination frequencies can be different in the meioses of the two parents. Van Ooijen and Jansen (2013) describe in detail linkage analysis of a full-sib family of an outbreeding species, including the two-way pseudo-testcross strategy, where the recombination frequency is estimated for each parent separately. This can be followed by integration of the male and female maps using bridge loci that segregate in both parents. Van Ooijen and Jansen (2013) also explain that a multipoint approach is superior for SNP markers with two alleles because of the way it can handle missing information in the observations. Computer software is available for doing the analyses but I think that an understanding of the underlying genetics is desirable, especially for designing good experiments in the first place. We will look at linkage and QTL analysis of a full-sib family under tetrasomic inheritance in Chap. 8.

## Mapping Functions

In general terms a mapping function ( $F$ ) can be defined as  $m_{ij} = F(r_{ij})$  where  $m_{ij}$  is the map distance (in Morgans) between loci  $i$  and  $j$  and  $r_{ij}$  is the recombination fraction ( $0 \leq r \leq 0.5$ ), and  $F$  exists for all pairs of loci (genes) and is a continuous function (Sham 1998).

If one considers three loci  $A$ ,  $B$  and  $C$  in that order (Fig. 4.3), with recombination frequencies (recombination fraction)  $p$ ,  $q$  and  $r$  between  $AB$ ,  $BC$  and  $AC$ , then recombination between  $A$  and  $C$  occurs when there is a cross-over between  $A$  and  $B$  but not between  $B$  and  $C$  and vice versa. Independence of chiasmata, and hence cross-overs, implies:

$$\begin{aligned} r &= p(1 - q) + q(1 - p) \\ r &= p + q - 2pq \end{aligned}$$

This can be rewritten as

$$1 - 2r = 1 - 2p - 2q + 4pq = (1 - 2p)(1 - 2q)$$

An additive relationship is then obtained by taking logarithms, in fact natural logarithms ( $\ln$ ) to the base  $e$ . The number  $e$  is irrational and its value correct to five decimal places is 2.71828 (Causton 1977).

$$\ln(1 - 2r) = \ln(1 - 2p) + \ln(1 - 2q)$$

Multiplying throughout by  $-1/2$  so that  $r$  becomes approximately equal to  $m$  at small distances gives Haldane's map function:

$$m = -\frac{1}{2}\ln(1 - 2r); \text{ or } r = \frac{1}{2}(1 - e^{-2m}).$$

A more general relationship that can allow for interference is as follows (Sham 1998).

$r = p + q - 2cpq$ , where  $c$  has a value between 0 (complete interference) and 1 (none) and is defined as the coincidence, and  $(1 - c)$  is defined as the interference.

To derive a corresponding map function it is useful to use  $\theta$  as the symbol for recombination frequency and to explore  $\theta$  as a function [ $f(m)$ ] of map distance ( $m$ ), i.e.  $\theta = f(m)$ . One also has to use some calculus.

If the map distances corresponding to  $r$ ,  $p$  and  $q$  are  $m + d$ ,  $m$  and  $d$ , then

$$f(m + d) = f(m) + f(d) - 2cf(m)f(d)$$

so that

$$f(m + d) - f(m) = f(d) - 2cf(m)f(d)$$

Division by  $d$  gives:

$$[f(m + d) - f(m)]/d = f(d)/d - 2cf(m)f(d)/d$$

If over short distances the map distance  $m$  equals the recombination frequency  $\theta$ , then as  $d$  tends to zero and  $[f(m+d) - f(m)]/d$  tends to the differential coefficient  $d\theta/dm$ ,  $f(d)/d$  tends to one and we have the following differential equation:

$d\theta/dm = 1 - 2c_0f(m) = 1 - 2c_0\theta$ , where  $c_0$  is the marginal coincidence when  $d$  approaches zero. Hence:

$$dm = d\theta/(1 - 2c_0\theta)$$

and

$$m = \int d\theta/(1 - 2c_0\theta) = -\ln(1 - 2c_0\theta)/2c_0$$

When  $c_0 = 1$  we have the Haldane map function:  $m = -\ln(1 - 2\theta)/2$

When  $c_0 = 2\theta$  we have the Kosambi map function:

$$m = \int d\theta/(1 - 4\theta^2) = \frac{1}{4}\ln[(1 + 2\theta)/(1 - 2\theta)]$$

Switching back to  $r$  for the recombination frequency:

$$m = \frac{1}{4}\ln[(1 + 2r)/(1 - 2r)]; \text{ or } r = (e^{4m} - 1)/[2(e^{4m} + 1)]$$

Although not additive, this mapping function has the following desirable properties.

When  $r = 0, c = 0$  and  $r = p + q$  (complete interference)

When  $r = \frac{1}{2}, c = 1$  and  $r = p + q - 2pq$  (no interference)

In between the amount of interference decreases as  $r$  increases; in other words, the further the loci are apart. In plants, molecular data in *Brassica* indicates that two cross-overs are not expected closer together than 15 cM (Kearsey and Pooni 1996).

A simple example of the two mapping functions is as follows [ $r_{AC} = r_{AB} + r_{BC} - 2c(r_{AB} \times r_{BC})$ ], where it can be seen that the Haldane one is additive and the Kosambi one almost additive:

$c = 1$	Haldane	$c = 2r_{AC}$	Kosambi
$r_{AB} = 0.150$	0.178	$r_{AB} = 0.150$	0.155
$r_{BC} = 0.150$	0.178	$r_{BC} = 0.150$	0.155
$r_{AC} = 0.255$	0.356	$r_{AC} = 0.275$	0.3095

## Test for Linkage and Membership of Linkage Group

Since the 1980s, an abundance of molecular (DNA) genetic markers have become available in all economically important crops and hence so have densely populated linkage maps of all of their chromosomes. For example, in potato an Ultra High Density genetic map is available comprising 10,000 (AFLP) markers (<http://www.dpw.wageningen-ur.nl/uhd>). Thus most differences between alleles at genetic loci ( $A/a$ ,  $B/b$ ,  $C/c$  etc.) are now detected at the DNA level rather than through phenotypic differences, although it is mapping genes of economic importance on the linkage maps that is of real interest to plant breeders.

Computer programmes such as JoinMap (Van Ooijen 2006) provide the software for constructing genetic linkage maps and such packages can be used without understanding how they work. Nevertheless, it is worth appreciating that partitioning markers into linkage groups and ordering loci within them are not trivial tasks.

When considering many loci, one needs to partition them into linkage groups so that a marker segregates independently of markers in different linkage groups and shows a significant association with at least some of the other markers within its linkage group. JoinMap uses LOD values as the criterion for linkage and considers all pairs of loci. A threshold LOD value is set below which linkage is not considered significant with the consequence that a higher threshold results in more and smaller linkage groups. Ideally, the number of linkage groups should equal the basic chromosome number of the species. At any stage in the process there will be groups of markers assigned to linkage groups and free markers not assigned. If none of the free markers is significantly linked by LOD value to one of the existing groups (through at least one marker in that group), a new linkage group is created. Otherwise a free marker which does show linkage with an existing group is added to that group. In contrast, in TetraploidMap (Hackett et al. 2007) which was developed for autotetraploids, the criterion for linkage is the significance of Pearson's chi-square statistic for a two-way contingency table. This significance can be transformed to give a measure of distance between loci and the values for all pairs of loci can fed into a cluster analysis. Different clustering methods will give slightly different dendograms. The nearest-neighbour cluster analysis adds a marker to a cluster according to its distance to the closest marker in the cluster, but can combine large groups on the strength of one marker from each subgroup. Such 'chaining' can be avoided by comparing the dendrogram from nearest-neighbour cluster analysis with that from average linkage cluster analysis as the criterion for linkage. Further discussion can be found in the book by Van Ooijen and Jansen (2013).

## Ordering Loci in Linkage Group

For three loci  $A$ ,  $B$  and  $C$  with recombination frequencies (recombination fraction)  $p$ ,  $q$  and  $r$  between  $AB$ ,  $BC$  and  $AC$ , we saw above that in the absence of interference the following relationship holds.

$$r = p(1 - q) + q(1 - p)$$

The sum of adjacent recombination coefficients (**sar**) for the order  $A$ ,  $B$ ,  $C$  is  $p + q$ .

If the order was mistakenly believed to be  $B$ ,  $A$ ,  $C$  then the **sar** would be calculated as  $p + r$ .

However, we know from the true order that  $p + r = p + p(1 - q) + q(1 - p) = p + q + p(1 - 2q)$ .

As the maximum value of  $q$  is  $\frac{1}{2}$ , the following is true:  $p + r \geq p + q$ .

Hence the true order is the one with the minimum **sar** and this property holds for any number of loci; but there are two problems with this as a criterion for determining the correct order. Firstly, the best order for a given set of loci may not be the true order simply because of sampling error. Secondly, there are too many orders to examine once the number of markers gets large; the number of ways of ordering  $m$  markers being  $m!/2$ . When  $m = 3$  this is simply  $(3 \times 2 \times 1)/2 = 3$  ( $ABC$ ,  $ACB$  and  $BAC$ ), as we have seen; but with  $m = 10$  the number is 1,814,400 and with  $m = 100$ , a staggering  $4.7 \times 10^{157}$ . Various alternative strategies have been devised such as seriation, simulated annealing and branch and bound (Weir 1996), and more are discussed by Van Ooijen and Jansen (2013). The first two methods examine relatively small and relatively large numbers of possible orders, respectively, but neither guarantee finding the true order. In contrast, branch and bound, provides an efficient search which is guaranteed to find the best order, but on occasions this might require an examination of all orders. Criteria for evaluating the best order include the largest sum of LODs of adjacent segments of the linkage map, the goodness of fit of the pairwise recombination frequency estimates to the recombination frequencies for adjacent loci obtained from the estimated map order, the minimum **sar** mentioned above, the least negative log-likelihood value and the minimum residual sum of squares in weighted least squares analysis (Van Ooijen and Jansen 2013).

The method implemented in JoinMap by Stam (1993) is essentially a trial and error procedure with a sequential build-up, using the most informative markers first and a reshuffling of the map order around a newly placed marker to avoid local optima in search for global one. This is because an exhaustive search of the parameter space for maps with over 50 markers soon becomes prohibitive. In other words, JoinMap does a numerical search for the best fitting linear arrangement. Starting from a pair of loci in the middle of the map, the next locus chosen has maximum linkage with the current loci (sum of pairwise LODs). JoinMap uses multilocus map distances and the order that is accepted is the one that minimizes the sum of (weighted) squares of differences between observed and expected distances.

Thus with three loci  $A$ ,  $B$  and  $C$  we have three recombination frequencies,  $r_{AB}$ ,  $r_{BC}$ ,  $r_{AC}$ , and three possible orders,  $ABC$ ,  $ACB$  and  $BAC$ . The recombination frequencies can be transformed to map distances,  $d_{AB}$ ,  $d_{BC}$ ,  $d_{AC}$ , by an inverse mapping function, either Haldane's or Kosambi's. The mapping function which best models the actual level of crossover interference should give the best results (the smallest deviations in the model below), but the level may be unknown. If we use  $x$  and  $y$  to denote the expected distances between adjacent loci, we have three possibilities.

1.  $ABC$ :  $A$  to  $B$  is  $x$ ,  $B$  to  $C$  is  $y$  and  $A$  to  $C$  is  $x + y$
2.  $ACB$ :  $A$  to  $C$  is  $x$ ,  $C$  to  $B$  is  $y$  and  $A$  to  $B$  is  $x + y$
3.  $BAC$ :  $B$  to  $A$  is  $x$ ,  $A$  to  $C$  is  $y$  and  $B$  to  $C$  is  $x + y$

The model for the first order, for example, is:

$$\begin{aligned} d_{AB} &= x + e_{AB} \\ d_{BC} &= y + e_{BC} \\ d_{AC} &= x + y + e_{AC} \end{aligned}$$

where  $e_{AB}$ ,  $e_{BC}$  and  $e_{AC}$  are the chance deviations from the true values.

The measures of discrepancies for the three orders are:

$$\begin{aligned} D_1 &= (d_{AB} - x)^2 + (d_{BC} - y)^2 + (d_{AC} - x - y)^2 \\ D_2 &= (d_{AC} - x)^2 + (d_{BC} - y)^2 + (d_{AB} - x - y)^2 \\ D_3 &= (d_{AB} - x)^2 + (d_{AC} - y)^2 + (d_{BC} - x - y)^2 \end{aligned}$$

Map distances can be estimated by a standard least squares procedure. Differentiating  $D$  with respect to  $x$  and  $y$  and setting  $\partial D / \partial x = 0$  and  $\partial D / \partial y = 0$ , yields a set of linear equations in  $x$  and  $y$  which can be solved by standard procedures to give the best estimates of multilocus map distances,  $x$ ,  $y$  and  $x + y$ .

If the true order is  $ABC$ , then  $D_1$  will have the lowest value (residual sum of squares) as we can see from our simple example above where the Haldane distances were  $d_{AB} = 0.178$ ,  $d_{BC} = 0.178$  and  $d_{AC} = 0.356$ , and solving  $\partial D / \partial x = 0$  and  $\partial D / \partial y = 0$  for  $D_1$ ,  $D_2$  and  $D_3$  in turn gives the following values for  $x$  and  $y$ :

1.  $ABC$ :  $x = 0.178$ ,  $y = 0.178$  and  $D_1 = 0$
2.  $ACB$ :  $x = 0.237$ ,  $y = 0.059$  and  $D_2 = 0.042$
3.  $BAC$ :  $x = 0.059$ ,  $y = 0.237$  and  $D_3 = 0.042$

In practice there will usually be more than three loci, only some of the possible recombination frequencies will be available, and the recombination frequencies and hence map distances will not be equally accurate. The latter problem can be solved by weighting the individual terms of  $D$ ; for example, LODs can be used as weights [ $(x - d_{AB})^2$  would be multiplied by the LOD for  $r_{AB}$ , etc.]. Then the same procedure is used as in our simple example. Van Ooijen and Jansen (2013) report that better results were obtained in practice with the square of the LODs as weights and these are used in the software package JoinMap.

Map estimation can also be done by maximum likelihood. When more than two loci are considered the method is referred to as multipoint maximum likelihood. An example can be found in the book by Van Ooijen and Jansen (2013). The model under consideration is defined by the order of the loci and the recombination frequencies between adjacent loci. The likelihood is the joint probability of the observations given the model, an observation being the phenotype over all loci of an individual in the mapping population. It is the product of the probabilities of the individuals in the mapping population as the individuals are independent. In order to estimate the map, the values of the recombination frequencies that maximize the log-likelihood (and hence the likelihood) are found, and then converted into map distances. Different orders of loci have different likelihoods, the best being the one with the least negative log-likelihood value. When all genotype information is available, the multipoint maximum likelihood estimates of recombination frequencies are identical to the two-point estimates. With missing genotypes, explicit estimators for the unknown recombination frequencies can no longer be found, and estimates are obtained using the EM algorithm. This is done analytically in the software package MAPMAKER (Lander and Green 1987) and stochastically in JoinMap (Van Ooijen 2006). The estimates will be different from the two-point ones and more accurate because information from all loci is taken into account.

Finally, one very important point needs to be made about error detection. Mistakes in typing or recording genotypes cause problems in genetic maps and the problems increase as the maps become denser. They tend to introduce spurious recombination events and hence inflate the genetic map. The simplest approach to detecting errors is to search for apparent double cross-overs in short map lengths as they are very unlikely to be genuine. Computer software is available for doing this (Weir 1996). Also, in general, a higher number of missing observations leads to a poorer linkage map (poorer fit of regression model). Van Ooijen and Jansen (2013) discuss pre- and post-mapping checks that are highly desirable. I personally find graphical genotypes of marker alleles for each linkage group very useful. These are displayed in colour or shades of grey according to parental source, in a two dimensional data matrix of markers in linkage map order (rows) and individuals (columns). Thus for any individual one can see where crossovers have occurred and whether or not they make sense.

## High Density SNP Maps

Next-Generation Sequencing of DNA is leading to the discovery of a large number of Single Nucleotide Polymorphisms (SNPs). It therefore seems appropriate to finish this chapter by looking at high density SNP maps. Let's consider an example from tomato ( $2n = 2x = 24$ ). Sim et al. (2012) developed a genotyping array of 7720 SNPs which had been discovered primarily from Next-Generation Sequencing of the whole genomes of six tomato germplasm accessions. In addition, 501 functional SNPs on the array were derived from candidate genes for traits such as disease

resistance and carotenoid biosynthesis. The array and its associated technology allowed the parents and offspring in three  $F_2$  mapping populations to be screened for the SNPs; in other words, high throughput genotyping was possible. As a result, high density linkage maps were constructed which could be compared with the physical maps of the genomes. The three interspecific populations were: EXPEN 2000 (*Solanum lycopersicum* LA0925  $\times$  *S. pennellii* LA0716, 79 individuals), EXPEN 2012 (*S. lycopersicum* Moneymaker  $\times$  *S. pennellii* LA0716, 160 individuals) and EXPIM 2012 (*S. lycopersicum* Moneymaker  $\times$  *S. pimpinellifolium* LA0121, 183 individuals). The numbers of markers segregating were 3503, 3687 and 4491, respectively. However, because of the relatively small population sizes, and hence recombination events, the numbers of unique map positions (known as genetic bins or BINS) were less, namely 1076, 1229 and 1358, respectively; with average BIN intervals of 1.6, 0.9 and 0.8 cM. All three linkage maps revealed an uneven distribution of markers across the genome. The 2000 map was considered inferior to the 2012 maps because of its inflated length due to the impact of order errors in a small population. Hence analysis concentrated on the two better maps which showed high levels of colinearity of markers across all 12 chromosomes, but also revealed evidence of small inversions of sequence between LA0716 and LA0121. The physical positions of 7666 SNPs were identified, using flanking sequences, relative to the tomato reference genome sequence available through the Solanaceae Genome Network (SGN; <http://solgenomics.net>). They covered a total of 758 Mb (out of predicted 900 Mb) of the tomato genome with an average density between markers of 0.12 Mb. Among the 7666 SNPs with physical positions 5296 SNP markers were mapped on one or both of the EXPEN 2012 and EXPIM 2012 genetic linkage maps. The genetic and physical positions were mostly consistent, as expected, with the vast majority (99.7 %) of SNPs in the linkage maps showing conserved chromosome assignments with the corresponding physical positions; the few exceptions were assumed to be errors in either the genetic or physical maps. Comparing genetic positions relative to physical positions revealed that genomic regions with high recombination rates were consistent with the known distribution of gene-rich euchromatin to the distal regions of all 12 chromosomes. The very low recombination rates were observed in the large pericentromeric regions within each chromosome, which represent repeat-rich and gene-poor heterochromatin encompassing 77 % of the tomato genome.

### ***Some Conclusions About High Density Maps***

The above example illustrates a number of important points. With a very large number of SNPs in a germplasm collection, one can expect to find a large number segregating in a particular cross and hence be able to achieve a dense linkage map. Furthermore, one can expect a reasonable number of SNPs to be segregating in more than one cross, and hence be able to compare the linkage maps from different crosses. Additionally, with dense linkage maps one can expect to find SNPs tightly

linked to genes of economic importance, or even SNPs within those genes, for use in marker-assisted selection. If a simply inherited trait is difficult or expensive to assess, or can only be assessed in a mature plant, then selection of seedlings having the marker can be an attractive proposition in a breeding programme. Equally, for unknown genes, one may be able to identify a candidate gene in the genome sequence close to the SNP marker. However, the frequencies of combinations of (desirable) alleles from different loci that arise in breeding and genetic experiments are determined by what happens during meioses. In other words, the frequencies are determined by the number of chromosomes in the species and the numbers and distribution of chiasmata on those chromosomes. It follows that the actual numbers will be determined by the population size. So the linkage map really is of importance in planning a breeding programme because it tells you what to expect.

# **Chapter 5**

## **Gene Expression and Selection of Major Genes**

### **Introduction**

In the last two chapters we explored the mapping of genes and how map positions affect the frequencies of desirable combinations of genes in breeding programmes. We now need to look at gene expression in some detail to see how these desirable combinations can be recognized and selected by plant breeders.

### **Biochemical Pathways**

Bateson et al. (1905) made a cross between two pure-breeding white-flowered varieties of sweet pea (*Lathyrus odoratus*) and found that all of the F<sub>1</sub> progeny had purple flowers. These were allowed to self-pollinate to produce the F<sub>2</sub> generation in which 382 plants had coloured petals and 269 were white; numbers which fit a 9:7 ratio. This ratio can be explained by two independent but complementary dominant factors so that the last three phenotypic classes in the 9:3:3:1 ratio for independent assortment are indistinguishable. Further breeding tests confirmed that one parental white-flowered strain had the genotype *CCrr* and the other had the genotype *ccRR*, where *C* and *R* are the complementary dominant genes required for anthocyanin pigment formation. One can envisage two successive biochemical steps, the end product of one being colourless and forming the substrate of the other. If these steps were controlled by the two genes, respectively, it would be necessary for both reactions to occur for pigment formation, and hence for both *C* and *R* to be present.

Bateson et al. (1905) also crossed a pure-breeding pink-flowered strain of *Salvia horminum* with a pure-breeding white-flowered one and found all of the F<sub>1</sub> progeny had purple flowers. The F<sub>2</sub> generation comprised 255 purple, 92 pink and 114 white-flowered plants, numbers which fit the 9:3:4 ratio expected if the last

two classes in the 9:3:3:1 ratio are indistinguishable. This explanation was confirmed by further breeding. Thus flower colour was determined by two independently-inherited factors such that when both dominant genes were present a purple pigment was produced, while one alone (*A*) gave a pink pigment, and the other alone (*B*) a white colour which was also found in the absence of either dominant allele (*aabb*). Again one can envisage the two genes acting sequentially in the synthesis of the purple pigment, gene *A* leading to a pink anthocyanin and gene *B* modifying its molecular structure to give a purple anthocyanin. In the absence of gene *A* no anthocyanin is formed and gene *B* is then ineffective. Bateson (1907) proposed the term epistatic for gene *A* and hypostatic for gene *B* because *A* 'stands above' *B* and determines whether or not it takes effect.

Subsequently biochemical explanations for these and similar results were confirmed and explained in terms of the genetic control, through enzymes, of the synthesis of flower pigments (Whitehouse 1973). Later in this chapter we will look at explanations of pigment colour for potato tubers. However, it was the pioneering work of Beadle and Tatum (1941) on nutritional mutants in fungi (*Neurospora*) that led to the theory of 'one gene: one enzyme'. It was found that specific genes blocked specific steps in the biosynthesis of arginine and tryptophan and that different mutants would respond to different precursors in the synthesis of these amino-acids. Beadle (1945) proposed that every biochemical reaction had a specific gene directing its course and that the function of genes was in directing the configurations of protein molecules, as all known enzymes are proteins.

## Epistasis and Modified Mendelian Ratios

The term epistasis is now used for all non-allelic interactions between different genes and numerous examples of the occurrence of modified Mendelian ratios are known. Some common examples of modified F<sub>2</sub> ratios are given below.

Genotypes	<i>A-B-</i>	<i>A-bb</i>	<i>aaB-</i>	<i>aabb</i>
Classical ratio	9	3	3	1
Dominant epistasis ( <i>A-</i> same)	---	12 ---	3	1
Recessive epistasis ( <i>aa</i> same)	9	3	---	4 ---
Duplicate genes with cumulative effect	9	---	6 ---	1
Duplicate dominant genes ( <i>A-</i> or <i>B-</i> )	-----	15 -----	-----	1
Duplicate recessive genes ( <i>aa</i> or <i>bb</i> )	9	-----	7 -----	-----
Dominant and recessive interaction	---	13 ----	3	---

'Duplicate recessive genes' is sometimes called complementary epistasis because *A* and *B* complement each other to produce a different phenotype.

The simple but important message for plant breeders is that they will often need to bring together a combination of alleles at different loci to achieve their objectives. Cyanogenesis provides a classic example.

## Cyanogenesis

Cyanogenesis is the release of cyanide following tissue damage and is common in plant species. In white clover (*Trifolium repens*) it provides one of the longest-studied and best-documented examples of an adaptive polymorphism in plants, with both cyanogenic and acyanogenic forms occurring. Their selective advantage and hence frequency is mainly but not exclusively determined by the presence of herbivores, as cyanogenesis provides resistance but carries an energetic cost to the plant. The polymorphism is controlled by two independently segregating genes: *Ac/ac* controls the presence/absence of cyanogenic glucosides; and *Li/li* controls the presence/absence of their hydrolyzing enzyme, linamarase. Hence both *Ac* and *Li* are required for cyanogenesis and we have a classical example of complementary epistasis. Olsen et al. (2008) have characterized *Ac* at the molecular level. *Ac* corresponds to a gene encoding a cytochrome P450 of the CYP79D protein subfamily (CYP79D15) which catalyzes the first step in cyanogenic glucoside biosynthesis. Southern hybridizations indicated that CYP79D15 occurs as a single-copy gene in cyanogenic plants but is absent from the genomes of *ac* plants. Gene-expression analyses by RT-PCR (reverse transcription–polymerase chain reaction) corroborated this finding. Likewise the *Li/li* polymorphism also arises through the presence/absence of a single-copy gene, with *Li* gene expression absent in plants lacking enzyme activity (Olsen et al. 2007). The authors suggest that the nature of these polymorphisms may reflect white clover's evolutionary origin as an allotetraploid derived from cyanogenic and acyanogenic diploid progenitors. We will encounter cyanogenesis again when we come to discuss cassava and white clover breeding.

## The Genetic Code

Dounce (1952) and Gamow (1954) independently had the idea that the linear sequence of nucleotides in nucleic acids was responsible for determining the linear sequence of amino-acids in the polypeptide chains of protein molecules. By 1973, in his book *Towards an Understanding of the Mechanism of Heredity*, Whitehouse (1973) was able to present extensive experimental evidence to demonstrate that nucleotide sequence specifies amino-acid sequence; indeed that there is an exact point-by-point relationship between the position of the mutation in the gene and the position of the amino-acid substitution in the polypeptide. Furthermore, by 1973 it was known how the sequence of four different bases in DNA was transcribed and translated into the sequence of 20 different kinds of amino acids in the polypeptide via messenger (m) RNA and transfer (t) RNA (ribonucleic acid). The genetic code had been cracked by an ingenious set of experiments that revealed consecutive non-overlapping triplets of nucleotides as the amino-acid code (Table 5.1 and Fig. 5.1).

**Table 5.1** The genetic (amino-acid) code with 5'-phosphates to the left and 3'-phosphates to the right of the symbols (A, C, G and U) for the ribonucleosides of adenine, cytosine, guanine and uracil, where AUG is also the polypeptide initiation codon and UAA, UAG and UGA are the stop signal codons

UUU	Phenylalanine	UCU	Serine	UAU	Tyrosine	UGU	Cysteine
UUC		UCC		UAC		UGC	
UUA	Leucine	UCA		UAA	Stop	UGA	Stop
UUG		UCG		UAG		UGG	Tryptophan
CUU		CCU	Proline	CAU	Histidine	CGU	Arginine
CUC		CCC		CAC		CGC	
CUA		CCA		CAA	Glutamine	CGA	
CUG		CCG		CAG		CGG	
AUU	Isoleucine	ACU	Threonine	AAU	Asparagine	AGU	Serine
AUC		ACC		AAC		AGC	
AUA		ACA		AAA	Lysine	AGA	Arginine
AUG	Methionine	ACG		AAG		AGG	
GUU	Valine	GCU	Alanine	GAU	Aspartic acid	GGU	Glycine
GUC		GCC		GAC		GGC	
GUА		GCA		GAA	Glutamic acid	GGA	
GUG		GCG		GAG		GGG	

DNA encoding strand	5'	ATG.....	ACCGGC.....	TAA	3'
DNA template strand	3'	TAC.....	TGGCCG.....	ATT	5'
RNA synthesis (transcription)					
mRNA	5'	AUG.....	ACCGGC.....	UAA	3'
		codon			
		anticodon			
tRNAs		UAC	UGG	CCG	
polypeptide synthesis (translation)		methion....threonine glycine.....STOP			
		START			

**Fig. 5.1** Transcription and translation

The information in the double-stranded helix of DNA is first transcribed (transcription) from the 3' to 5' strand (template strand) into single-stranded messenger RNA in the nucleus. Thus RNA synthesis proceeds from the 5' to the 3' end of the mRNA molecule. The structure of RNA is similar to DNA but differs in that the sugar is D-ribose instead of 2-deoxy-D-ribose, and one of the pyrimidines is uracil (U) in place of the thymine (T) (5-methyluracil). The nucleotides are polymerized by 3', 5'-phosphodiester linkages as in DNA to form a polynucleotide chain. The DNA strand complementary to the template strand is called the encoding strand

because it has the same sequence as the mRNA, apart from T in place of U. The mRNA moves through the pores in the nuclear membrane to the site of polypeptide synthesis in the cytoplasm, namely the ribosomes. Each amino-acid is first combined with its own specific adaptor (a transfer RNA) which associates it with a particular triplet of bases in the mRNA (the codon). A structural component of ribosomes is a third type of RNA referred to as ribosomal (r) RNA. The direction of translation of the nucleotide sequence in mRNA into the amino-acid sequence in a polypeptide is the same as the direction of transcription from DNA to RNA. The genetic code was established for the bacterium *Escherichia coli* *in vitro* (cell-free studies). What is remarkable is that the same code applies *in vivo* to virtually all organisms; it is universal (Whitehouse 1973).

## Herbicide Tolerant Crops

Herbicide tolerant crops provide a good example of understanding the biochemical and genetic basis of an economically important trait. Today high input farming relies on the use of herbicides to control weeds in crops. One weed management strategy is the use of herbicide tolerant crops in combination with their corresponding herbicides. It is particularly effective when the weed is closely related to the crop species. Herbicide tolerant crops can also be good rotational crops for fields where herbicide residue has carried over in the soil from previous years. A susceptible crop can be genetically changed to a tolerant one by mutation of a native gene, either spontaneous or induced, or by insertion of a foreign gene by transformation. As a result of the genetic change, either the enzyme encoded by the gene becomes insensitive to the herbicide or it catalyzes the degradation of the herbicide before it can become effective. Commercial crops developed from mutations include imidazolinone-tolerant maize, rice, wheat, rapeseed, sunflower and lentil; sulphonylurea-tolerant soybean and sunflower; cyclohexanedione-tolerant maize; and triazine-tolerant rapeseed (Tan and Bowe 2012). All of these are examples of genes that encode altered enzymes targeted by herbicides.

Imidazolinones and sulphonylureas inhibit a critical enzyme of amino acid biosynthesis, acetohydroxyacid synthase (AHAS), with nuclear coding gene *AHAS*. Five commonly occurring mutation sites have been found that confer tolerance to these herbicides in plants. For example, maize has two *AHAS* gene loci, one on the long arm of chromosome 4 and the other on the long arm of chromosome 5. Mutations have been found in both at the codon numbered 653 in which a single nucleotide substitution from AGU to AAU results in the amino acid substitution serine to asparagine in the AHAS enzyme. The amino acid substitution modifies the herbicide-binding pocket of the enzyme and prevents the herbicide from binding and inhibiting the enzyme. The same mutation has been found in the *AHAS* gene of rice, one (*AHAS1*) of the two active *AHAS* genes of rapeseed (there are three others), and two of the three *AHAS* genes of wheat. Higher levels of imidazolinone tolerance in wheat can be achieved by stacking two or more tolerant

genes into a single genotype. An even more effective mutation has been found in the other active rapeseed gene (*AHAS3*), and in maize, at codon 574 where the nucleotide substitution UGG to UUG results in the amino acid substitution tryptophan to leucine. *AHAS1* is tolerant to imidazolinones only, whereas *AHAS3* is cross tolerant to both imidazolinones and sulfonylureas. The tolerance of *AHAS3* to imidazolinones is greater than that of *AHAS1* but the highest level of tolerance comes from combining the two mutations with homozygosity at both loci. In rice another effective mutation has been found at codon 654, with the nucleotide substitution GGG to GAG resulting in the amino acid substitution glycine to glutamic acid. A list of characterized *AHAS* mutations can be found in the review by Tan and Bowe (2012). To date, tolerant alleles in commercial crops show incomplete dominance but no adverse agronomical effects.

Cyclohexanedione inhibits a key enzyme of fatty acid biosynthesis, acetyl-CoA carboxylase, with nuclear coding gene *ACCase*. Two commonly occurring mutation positions have been found, one of which in maize shows incomplete dominance with no adverse agronomical effects. Finally, triazines block electron transport in photosynthesis II by binding to the D1 protein with chloroplast coding gene *psbA*. Two commonly occurring mutation sites have been found, one of which has been used to develop triazine-tolerant rapeseed. In rapeseed the nucleotide substitution from AGU to GGU at codon 264 resulted in amino acid substitution serine to glycine. As a consequence, triazines can no longer bind to the D1 protein and interrupt electron transport in photosynthesis II. The trait is inherited maternally and tends to be associated with low yield and oil content and poor seedling vigour.

In summary, for commercially acceptable tolerance to herbicides, some herbicide-tolerant alleles can be heterozygous, others need to be homozygous, and the rest must be combined with another tolerant gene (Tan and Bowe 2012). Nevertheless, their simple inheritance means that the herbicide tolerance trait can easily be incorporated into an elite cultivar by a backcrossing programme. It is however important that herbicide-tolerant crops are properly managed by farmers through appropriate stewardship schemes in order to preserve their effectiveness as weed-management tools (Tan et al. 2005). Otherwise gene flow (outcossing) might occur to closely related weed species or spontaneous mutations to tolerance in weed species might be selected. After all, some of the spontaneous crop mutants were found as volunteers when a following crop of a different species was sprayed with herbicide. Two examples of good practice to reduce the risk of developing herbicide-resistant weeds are growing imidazolinone-tolerant crops for a maximum of 2 out of every 4 years in the same field and rotating herbicides with different modes of action in the same field. It also makes sense to use imidazolinone-tolerant maize and rice in rotation with soybeans as imidazolinones are common herbicide choices for soybeans. This eliminates any risk of maize or rice injury resulting from carryover of residual imidazolinone herbicides from the previous year in soybeans.

## Control of Gene Expression

Since the 1950s there have been tremendous advances in our understanding of the control of gene expression, and hence in our understanding of the temporal and spatial genetic control of plant development from a fertilized egg. In a plant containing between 25,000 and 40,000 protein-coding genes, most of these genes are transcribed only in certain organs and cells and only at specific times. Therefore the transcription of these genes is subject to complex and specific regulation. The genes for enzymes of metabolism and protein synthesis which proceed in all cells, independently of their specialization, are transcribed more often and are referred to as housekeeping genes. It was Jacob and Monod and their associates, working with *E. coli*, who in the 1960s recognized two kinds of genes: structural genes which determine the amino-acid sequence of specific proteins, and regulatory genes which control the functioning of the structural genes (Whitehouse 1973). However, control of gene expression in plants proved different to that in bacteria. The interested reader can refer to the second edition of *Plant Biotechnology, the genetic manipulation of plants* by Slater et al. (2008) or the fourth edition of *Plant Biochemistry* by Heldt and Piechulla (2011) for detailed information. For the purposes of this book it will suffice to understand the role of the promoter and transcription factors. However, the earlier description of transcription needs to be modified.

### Promoter

Each protein-coding gene is associated with its own regulatory regions, the most important of which is the promoter. The promoter contains the binding site for the enzyme RNA polymerase II, which makes an RNA (pre-mRNA) copy of the template strand of DNA during transcription. As both the complementary encoding strand of DNA and the single strand of mRNA are read from the 5' to 3' end of the polynucleotide, the promoter is said to be at the 5' end of the gene. The DNA sequences which occur before the transcription starting point are referred to as upstream and those that occur afterwards as downstream. Thus the promoter element is found about 25 bp upstream from the transcription start site. The coding regions of the pre-mRNA are the exons which carry the information translated into amino-acid order during protein synthesis. Non-coding regions (introns) between exons are removed (spliced out) during pre-mRNA processing, which also includes the addition of a G-cap sequence to the 5' end and a poly (A) sequence of up to 250 bp to the 3' end near a poly (A) addition signal. Exon/intron boundaries are delineated by specific base sequences that provide signals for cleavage and removal of the intron by riboprotein complexes. The resulting mRNA is then exported from the nucleus to the cytoplasm where it associates with ribosomes and is translated in the 5' to 3' direction to produce the protein product.

The promoter contains various sequence elements that interact with proteins that regulate transcription of the gene. For example, the TATA (25 bp upstream) and CAAT (80–110 bp upstream) boxes are involved in positioning and enhancing the activity of RNA polymerase II. Although the DNA sequences of promoter elements vary between genes and between species, there are consensus sequences in which each nucleotide is found in the majority of the sequences. Housekeeping genes often contain a C-rich region instead of the CAAT box. Sometimes 1000 bp upstream, several sequences can be present which function as (*cis*-regulatory) enhancers or silencers of gene expression.

## ***Transcription Factors***

Other sequence elements are involved in binding specific proteins which are required for transcription by RNA polymerase. These proteins are called transcription factors and link various signals (external and internal) to gene expression. One of their key features is their DNA-binding domain and hence they are also referred to as sequence-specific DNA-binding factors. Basal factors are essential for transcription whereas *trans*-elements, activators and repressors, enhance or slow down the process. Transcription factors are therefore important in determining the level, place and timing of gene expression. Similar sequence elements are commonly found in genes with similar functions or which respond to similar signals. One type of transcription factor contains three to nine zinc fingers, each of which can bind a sequence of three nucleotides on the DNA. Each finger consists of a peptide chain with two cysteine residues, two to four amino acids apart, separated by 12 amino acids from two histidine residues, three to four amino acids apart. A zinc ion is bound between the cysteine and histidine residues. Another type of transcription factor, called a leucine zipper, is a dimer of DNA binding proteins, where each monomer comprises a DNA binding domain of basic amino acids and  $\alpha$ -helix with three to nine leucine residues. The helices are held together by hydrophobic interactions between the leucine residues.

Pérez-Rodríguez et al. (2010) maintain a plant transcription factor database (<http://plntfdb.bio.uni-potsdam.de/v3.0/>) that provides putatively complete sets of transcription factors (TFs) and other transcriptional regulators (TRs) in plant species (*sensu lato*) whose genomes have been completely sequenced and annotated. The complete sets of 84 families of TFs and TRs from 19 species ranging from unicellular red and green algae to angiosperms are included in the database. For example, it is estimated that the *Arabidopsis thaliana* genome encodes 1500 factors for the regulation of gene expression (Riechmann et al. 2000). Clearly there are other opportunities for the control of gene expression, from access to DNA in chromosomes, through processing, transport and degradation of mRNA, to post-translational modification of proteins. Small (sm) RNAs consisting of 21–24 nucleotides, for example, regulate the expression of genes involved in responses to stress and lack of nutrients (Heldt and Piechulla 2011). Several smRNA types

bind to complementary sequences of target mRNAs to form a double-stranded (ds) RNA. This leads to the degradation of the mRNA by RNase II and other degrading enzymes and an inhibition of translation. This property is utilized in biotechnology to suppress the expression of a targeted gene by RNA interference (RNAi). Plants can be made to synthesize a small RNA complementary to a defined mRNA by genetic transformation. The technique can also be used to identify gene function by inhibiting the expression of a gene and evaluating the effects on the plant.

We are now in a position to examine in more detail some of the genes that were important for establishing the laws of inheritance, for domestication and adaptation of crops to new environments, for selection of colour and shape and for the green revolution. They are all simply inherited genes with phenotypes that were easy to recognize and select by farmers and plant breeders. Today, however, we know a lot more about the underlying physiology and biochemistry. This gives us insight into why some major genes, but not others, have been of great value to plant breeders; and should provide a more rational basis for mutation breeding.

## Mendel's Genes

Mendel determined the laws of heredity by using seven pairs of traits in garden pea: cotyledon colour, seed shape, pod colour, pod shape, flower colour, flower position and stem length. Seed shape (smooth versus wrinkled) and stem length are regulated by the *R* and *Le* loci and correspond to natural mutations of a starch-branching enzyme gene and a gibberellin 3 $\beta$ -hydroxylase gene, respectively.

Mutant *r* at the *R* (rugosus) locus has clearly visible effects on seed phenotype as a result of its profound effects on the composition of the developing pea seed. As well as being wrinkled, *rr* seeds contain smaller amounts of starch than *RR* seeds and have much lower ratios of branched amylopectin to amylose, indicative of a problem with its starch-branching enzyme. Bhattacharyya et al. (1990) cloned and characterized the gene (cDNA) for starch-branching enzyme and found that wrinkled (*rr*) seeds lacked one isoform of starch-branching enzyme (*SBEI*) present in round (*RR* or *Rr*) seeds. A major polymorphism in the *SBEI* gene between near-isogenic *RR* and *rr* lines showed 100 % cosegregation with the seed phenotype in 79 F<sub>6</sub> plants from two *RR* × *rr* crosses. An aberrant transcript for *SBEI* was produced in *rr* embryos. In *rr* lines the *SBEI* gene was interrupted by a 0.8 kb insertion that was very similar to the *Ac/Ds* family of transposable elements from maize. Failure to produce *SBEI* had complex metabolic consequences on starch, lipid, and protein biosynthesis in the seed.

Martin et al. (1997) provided molecular evidence that locus *Le* encodes the gibberellin (GA) 3 $\beta$ -hydroxylase that converts GA<sub>9</sub> to GA<sub>4</sub> and GA<sub>20</sub> to GA<sub>1</sub>, the latter being the major gibberellin controlling stem elongation and hence stem length in pea. They isolated cDNAs (transcribed genes) encoding a GA 3 $\beta$ -hydroxylase from pea lines carrying the *Le* allele and three mutant alleles *le*, *le-3* and *le<sup>d</sup>* that

result in dwarfism. The mutant alleles *le* and *le-3* each contained a base substitution resulting in single amino acid changes relative to the sequence from *Le*. Expression products from *le* and *le-3* possessed similar levels of 3 $\beta$ -hydroxylase activity for both substrates, GA<sub>9</sub> and GA<sub>20</sub>, that was less than the activity from *Le*. The allele from *le<sup>d</sup>*, a mutant derived from a *le* line, contained both the *le* mutation and a single-base deletion, which caused a shift in reading frame and presumably a null mutation. As expected, the expression product from *le<sup>d</sup>* was inactive. The *Le* gene was also isolated and studied by Lester et al. (1997).

Cotyledon colour (yellow versus green) is regulated by locus *I* with mutant *i* retaining not only greenness of the cotyledon during seed maturation but also greenness of leaves during senescence, suggesting that this mutant is a stay-green mutant. Sato et al. (2007) found tight linkage between the *I* locus and the stay-green gene originally found in rice, *SGR*. Molecular analysis of three *i* alleles including one with no *SGR* expression confirmed that the *I* gene encodes *SGR* protein in pea. Hence the *I* gene should probably be renamed *PsSGR*. Sato et al. (2007) suggest that *SGR* might be involved in activation of the chlorophyll-degrading pathway during leaf senescence through translational or post translational regulation of chlorophyll-degrading enzymes.

Mendel's fourth gene that is understood at the molecular level is his flower colour gene (*A*) which is a (bHLH) transcription factor that controls the anthocyanin biosynthesis pathway. When mutated, the transcription factor becomes inactive, anthocyanin is not produced, and the result is white rather than violet-red flowers. Hellens et al. (2010) used the genome sequence of model legumes, together with their known synteny to the pea genome, to identify a candidate gene for the *A* locus in pea. They then used molecular genetics analyses to show that the white flowered mutant allele, most likely used by Mendel, is a simple G to A transition in a splice donor site that leads to a mis-spliced mRNA with a premature stop codon.

## Domestication Genes

In an article in *Science* on how ancient farmers turned weeds into crops, Doebley (2006) mentions six domestication genes, of which five encode transcription factors that regulate other target genes by directly binding to their DNA. Hence they can have a major influence in controlling plant development. Importantly, these domesticated alleles are functional, not null or loss-of-function alleles, and this implies that domestication has involved a mix of changes in protein function and gene expression to adapt plants to cultivation. Doebley (2006) raises the prospect of modern plant breeders tinkering with the expression patterns or protein functions of known domestication genes to create superior alleles. The reader may also find useful the review by Martínez-Andújar et al. (2012) on seed traits modified during domestication: size, yield, dispersal and shattering, and dormancy and germination. So let us take a closer look at some domestication genes.

## **Maize**

Maize (*Zea mays*) was domesticated from annual teosinte (*Zea mays* subspecies *parviglumis*) around 9000 years ago in the mid to lowland regions of southwest Mexico where subspecies *parviglumis* grows endemically (van Heerwaarden et al. 2011). The domestication of maize provides a striking example of an increase in apical dominance in a crop plant compared with its wild ancestor. Branches (tillers) initiated at the base of grass plants are suppressed in response to shade conditions. This suppression of tiller and lateral branch growth was an important trait selected by early agriculturalists during maize domestication and crop improvement. Doebley et al. (1997) used transposon tagging to clone the *teosinte branched1* (*tb1*) gene which had been identified as a major contributor (QTL) to this change. They provided evidence that *tb1* acts both to repress the growth of axillary organs and to enable the formation of female inflorescences. They found that the maize allele of *tb1* is expressed at twice the level of the teosinte allele, suggesting that gene regulatory changes underlie the evolutionary divergence of maize from teosinte. Whipple et al. (2011) cloned another gene, *grassy tillers1* (*gt1*), that is dependent on the activity of *tb1* and whose expression is induced by shading. They found that it encodes a class I homeodomain leucine zipper gene that promotes lateral bud dormancy and suppresses elongation of lateral ear branches. Like *tb1*, *gt1* maps to a quantitative trait locus that regulates tillering and lateral branching. An even more important step in the domestication of maize was the liberation of the kernel from the hardened, protective casing that envelops it in teosinte. This exposed the kernel on the surface of the ear so that it could readily be used by humans as a food source. Wang et al. (2005) found that this change was controlled by a single gene (*teosinte glume architecture* or *tga1*) belonging to the SBP-domain family of transcriptional regulators, with just seven differences in DNA sequence between maize and teosinte: one difference encoded a non-conservative amino acid substitution that may affect protein function and the other six potentially affected gene regulation.

## **Rice**

Rice (*Oryza sativa*) was domesticated from the annual type (*O. nivara*) of wild rice (*O. rufipogon*). A key domestication trait was a non seed-shattering habit. Li et al. (2006) analyzed an F<sub>2</sub> population from crossing *O. sativa* ssp. *indica* and *O. nivara* and identified a QTL allele of large effect for the reduction of grain shattering in cultivated rice. The *sh4* allele (*shattering4*) of the wild species caused shattering, was dominant, and mapped to chromosome 4. Accessions of *O. nivara*, *O. rufipogon* and the other wild A-genome species, with confirmed shattering, differed from cultivars of *O. sativa* by one mutation, which was a nucleotide substitution of G for T that resulted in an amino acid substitution of asparagine

(neutral) for lysine (positively charged) in *O. sativa* in a predicted DNA binding domain, presumably of a transcription factor. Transformation of *O. sativa* ssp. *japonica* cultivar Taipei 309 with constructs that differed at the mutation site provided confirmation that the amino acid substitution was primarily responsible for the reduction of grain shattering in rice domestication. The substitution undermined the gene function necessary for the normal development of an abscission layer that controls the separation of a grain from its pedicel. Shattering was severely weakened but not eliminated so that the grains were retained on the plants long enough for harvest, but could then be removed by threshing. A second major shattering QTL (*qSH1*) of similar effect was found by Konishi et al. (2006) in a cross between ssp. *indica* and ssp. *japonica* rice.

Another key domestication trait was straw-white panicles filled with ripened grains in contrast to the black hull of wild rice which was thought to be a natural colour for shattered seeds in a dark mud land. Zhu et al. (2011) cloned (map-based cloning) a *Bh4* gene encoding an amino acid transporter protein and found that a 22-bp deletion within the third exon caused the loss of function of *Bh4* (frame shift gave stop codon which truncated the protein by 150 amino acids), resulting in the transition from black hull in wild rice to straw-white hull in cultivated rice. The 22-bp deletion accounted for 94.9 % of the straw-white hull colour in the screened cultivated rice varieties. A transgenic study confirmed that the *Bh4* gene controlled the synthesis of black hull pigment during seed maturation. Although the prevalent mutation of *Bh4* was the 22-bp deletion within the third exon, there were other mutations in this gene that led to the loss of black hull pigment. This suggested that different mutations occurred in the *Bh4* gene during rice domestication and that this phenotype of cultivated rice might have multiple origins. When the *Bh4* gene was transferred into the two subspecies of cultivated rice, *indica* (Guangluai 4, Kasalath) and *japonica* (Nipponbare), only the *indica* varieties gained the black hull phenotype. Hence some other genes in the pigmentation pathway must have mutated in *japonica* varieties during domestication.

## **Sorghum**

Tang et al. (2013) chose sorghum as a favourable species to investigate the genetic basis of shattering because only one genetic locus, *Sh1*, explains 100 % of the phenotypic variance in a cross between an elite *Sorghum bicolor* breeding line (BTx623), and wild *S. propinquum*. Furthermore, *S. bicolor* has a modest genome size of about 730 Mb (twice that of rice), a high quality reference sequence and shows similar composition and high levels of synteny and microcolinearity with maize and rice despite about 50 million years of divergence. Tang et al. (2013) used molecular genetic analysis to implicate a WRKY transcription factor, SpWRKY, in conferring shattering to *S. propinquum*. For proof of function they transformed recessive nonshattering sorghum cultivar Tx430 with genetically dominant SpWRKY and observed discernible seed dropping. They found that the SpWRKY

protein is 44 amino acid residues longer than the SbWRKY one because an AUG to AUU mutation resulted in a methionine (start codon) to isoleucine substitution. They proposed that SpWRKY functions in a manner analogous to *Medicago* and *Arabidopsis* homologues that regulate cell wall biosynthesis genes, with low expression toward the end of floral development derepressing downstream cell wall biosynthesis genes to allow deposition of lignin that initiates the abscission zone in the seed–pedicel junction. In contrast, the high expression of SbWRKY in the final floral development stage may keep cell wall biosynthesis genes repressed in *S. bicolor*, also repressing the shattering phenotype. Interestingly, the shattering allele of SpWRKY appears to be a gain-of-function mutation in *S. propinquum* after its divergence from a common ancestor shared with *S. bicolor*, and hence it played little or no role in the domestication of sorghum. Rather, it reinforced or further increased the level of shattering in *S. propinquum* resulting from another transcription factor, YABBY. The YABBY locus lies only 300 kb from the SpWRKY locus. The YABBY transcription factor contributes to seed shattering in crosses between a shattering race, Virgatum, and a nonshattering breeding line Tx430 of *S. bicolor*, and presumably in the common ancestor just mentioned. During domestication of wild *S. bicolor* at least three independent mutations occurred in the YABBY transcription factor and were selected by farmers, the outcome being non-shattering cultivated *S. bicolor*.

## Wheat

Simons et al. (2006) did a molecular characterization of gene *Q* on chromosome 5A, whose mutation from allele *q* to *Q* in an ancestor of modern wheats was largely responsible for the widespread cultivation of wheat. It confers the free threshing character and pleiotropically affects many other domestication-related traits such as glume shape and tenacity, rachis fragility, spike length, plant height and spike emergence time. It allowed early farmers to efficiently harvest the grain and today in modern agriculture is essential for the mass production and mechanical harvesting of the crop. Simons et al. (2006) were able to show that *Q* is a member of the AP2 class of transcription factors, which are characterized by having two plant-specific DNA binding motifs. They isolated the *Q* gene and verified its identity by analysis of knockout mutants, produced by chemical treatment with ethyl methanesulfonate (EMS), and by transformation. *Q* most likely arose through a gain of function mutation as both *Q* and *q* alleles are functional, but confer different phenotypes. Although several bases at the DNA level defined structural differences between *Q* and *q*, only a single putative amino acid at position 329 differentiated the *Q* and *q* proteins. Isoleucine in place of valine in the *Q* protein was thought to be important for homodimer formation and an increased level of transcription. The mutation that gave rise to the *Q* allele probably occurred only once, but it is not clear if it first arose in tetraploid or hexaploid wheat.

Two more major threshability genes have been analyzed by Sood et al. (2009). The wild wheat progenitors had tough glumes enveloping the floret that also made spikes difficult to thresh, whereas cultivated wheats have soft glumes. In hexaploid wheat, the glume tenacity gene *Tg*, as well as the major domestication locus *Q*, controls threshability. Sood et al. (2009) used comparative mapping to demonstrate that the soft glume (*sog*) gene of diploid *Triticum monococcum* and the tenacious glume (*Tg*) gene of hexaploid *T. aestivum* are not true orthologues (non-syntenic positions on chromosomes 2AS and 2DS) and hence could have arisen independently.

## **Barley**

In their paper on the genetic basis of six-rowed barley (*Hordeum vulgare* ssp. *vulgare*), Komatsuda et al. (2007) explain why this characteristic should be regarded as a domestication trait, along with nonbrittle rachis. Wild barley (*H. vulgare* ssp. *spontaneum*), the progenitor of cultivated barley, is two-rowed; its arrow-like triple spikelets helping the seeds to bypass stones and pebbles and reach soil when they fall to the ground. In contrast, spontaneous six-rowed mutants are at a disadvantage in wild barley populations because they lack this adaptation. Archaeological specimens from the Near East have revealed two-rowed barley with a brittle rachis in pre-agricultural sites dating from 19,000 to 9000 years ago, two-rowed barley as the earliest domesticated type dating from 9500 to 8400 years ago, and the start of cultivation of six-rowed barley around 8800–8000 years ago. Six-rowed barley became dominant from 7000 to 6000 years ago, presumably because of its higher yield. All three spikelets of modern six-rowed barley cultivars are fully fertile and able to develop into grains whereas the lateral spikelets of two-rowed barley are reduced in size and are sterile. The development of a six-rowed spike is controlled by a single allele *vrs1* that is recessive to the dominant allele responsible for the two-rowed spike *Vrs1*, the gene being on chromosome 2H. Komatsuda et al. (2007) isolated *vrs1* by means of positional cloning, which involved developing a high-resolution linkage map and a physical map, comprising six key bacterial artificial chromosome (BAC) clones which were sequenced. In other words, the section of barley chromosome containing the gene was cloned in bacteria. Komatsuda et al. (2007) demonstrated that the wild-type *Vrs1* allele encodes a transcription factor that includes a homeodomain with a closely linked leucine zipper motif. Expression of *Vrs1* was strictly localized in the lateral-spikelet primordia of immature spikes, suggesting that the VRS1 protein suppresses development of the lateral rows. The *Vrs1* region was completely deleted in seven mutations produced by irradiation of two-rowed barley. Loss of function of *Vrs1* resulted in complete conversion of the rudimentary lateral spikelets in two-rowed barley into fully developed fertile spikelets in the six-rowed phenotype. Phylogenetic analysis demonstrated that the six-rowed phenotype originated repeatedly, at different times and in different regions, through independent

mutations of *Vrs1*. When five naturally occurring alleles were compared, the two- and six-rowed barleys showed equal levels of transcription of *Vrs1*, thus single-base-pair mutations in *vrs1* appear to be responsible for the functional changes observed in all three six-rowed alleles. Around 1000 years ago, two-rowed barley was introduced from the Near East to Europe and the allele *Vrs1.b3* has expanded rapidly throughout the world and become dominant. Today high yielding two-rowed as well as six-rowed barley cultivars are available. The size of the lateral spikelets is controlled by the *Int-c* locus on chromosome 4H, with large size dominant (*Int-c/Int-c*) over small size (*int-c/int-c*) (Horsley et al. 2009). Hence the genotype of a six-rowed spike is *vrs1vrs1Int-cInt-c* and the genotype of a two-rowed one is *Vrs1Vrs1int-cint-c*. There are no problems in making crosses between six-rowed and two-rowed barley and the parental genotypes are recovered as expected.

### **Number of Domestication Genes**

We have looked at domestication traits, examined their genetics, and found relatively few domestication genes. However, with the advent of Next-Generation Sequencing, another approach can be taken to answer the question, how many domestication genes? As an example, let's return to the paper of Huang et al. (2012b) considered earlier under domestication. Huang et al. (2012b) constructed a comprehensive map of rice genome variation from sequences from 446 geographically diverse accessions of the wild rice species *Oryza rufipogon*, the immediate ancestral progenitor of cultivated rice, and from 1083 cultivated *indica* and *japonica* cultivars. In a search for signatures of selection, they identified 55 selective sweeps that have occurred during domestication. These signatures result from a reduction in nucleotide variation (diversity) around the selected loci in the cultivated accessions compared with the wild ones ( $\pi_w/\pi_c$ ), as a result of tight linkage between the selected loci and their neighbours. Altered allele frequencies are also expected in the domestication loci. Most well-characterized domestication genes, such as *sh4* and *Bh4*, were among the 55 loci detected. In addition, QTL mapping (see next chapter) of 15 domestication-related traits was performed on a population of 271 lines developed from a cross between *O. sativa* subspecies *indica* Guangluai-4 and *O. rufipogon* (Or-IIIa accession W1943). Of 58 loci detected, 32 were within domestication sweeps and almost all the peak signals for ten QTLs were within 200 kb of known causal genes. Finally, for domestication loci with no known genes, Huang et al. (2012b) found a total of 273 novel functional variants within 204 genes that showed high differentiation in allele frequencies between cultivated and wild rice. Hence for rice we should probably think in terms of at least tens of domestication genes rather than just a few key ones. For maize the work of Hufford et al. (2012) suggests the number may be hundreds. They generated 781 Gb of sequence from 35 improved maize lines, 23 traditional landraces and 17 wild relatives.

## Adaptation to New Environments

### **Photoperiod: Wheat**

Wheat (*Triticum aestivum*) is a quantitative long-day plant, whose flowering is accelerated by long days. Photoperiod insensitive varieties flower under both short and long-day conditions whereas photoperiod sensitive types show delayed heading, or do not head at all, if the day length and number of long days do not reach a threshold for floral initiation. Wheat is also one of the most adaptable of crops and is planted in virtually all countries over which the photoperiod varies dramatically and continually, suggesting there may be corresponding variations in photoperiod response among wheat varieties adapted to different environments. The genetic control of photoperiod sensitivity in wheat is primarily determined by the *Ppd-D1* and *Ppd-B1* genes located on the homoeologous chromosomes 2D and 2B, respectively, where the photoperiod insensitive alleles are *Ppd-D1a* and *Ppd-B1a*, and the sensitive ones are *Ppd-D1b* and *Ppd-B1b*, respectively. Guo et al. (2010) used molecular markers to score a total of 547 accessions of wheat for allelic variation at the *Ppd-D1* locus. The accessions comprised 492 common wheat cultivars (genomes AABBDD) selected from 41 countries on six continents, 25 synthetic wheats derived from crosses of tetraploid wheat (AABB) and *Aegilops tauschii* (DD), and 30 *A. tauschii* accessions. Guo et al. (2010) found six alleles (haplotypes) which were differently distributed among common wheat and *A. tauschii*. Only haplotypes II, V and VI were present in *A. tauschii*, whereas haplotypes I, II, III and IV were present in common wheat, indicating that haplotypes V and VI were relatively ancient and unselected. The distribution frequency of the haplotypes showed partial correlations with both latitudes and altitudes of wheat cultivation regions. Hence their results confirm the presence of more genetic variation for photoperiod response than once thought.

### **Photoperiod: Maize**

Maize (*Zea mays* L. subsp. *mays*) was domesticated from Mexican teosinte (*Zea mays* L. subsp. *parviglumis*), a species adapted to day lengths of less than 13 h (Hung et al. 2012a). The pre-Columbian spread of maize from its centre of origin in tropical Southern Mexico to the higher latitudes of the Americas required post domestication selection for adaptation to longer day lengths. Flowering time of teosinte and tropical maize is delayed under long day lengths, whereas temperate maize evolved a reduced sensitivity to photoperiod. Hung et al. (2012a) did extensive genetic analyses of appropriate temperate maize, tropical maize and teosinte germplasm to determine the genetic basis of flowering time and one of its components, photoperiod response. Photoperiod response was measured as the difference in mean thermal time (growing degree-days) to flowering between

long- and short-day length environments. Thermal rather than actual time was used to minimize the effect of temperature differences among environments on the observed flowering times. The results in brief are as follows. Photoperiod responses for thermal time to male and female flowering were highly correlated ( $r = 0.92$ ) but the latter had a higher heritability; hence the focus was on time to female flowering (i.e. silking). Photoperiod response was under moderately complex genetic control, regulated in one experiment by at least 14 Quantitative Trait Loci (QTLs) which were largely a subset of 29 QTLs for flowering time under long days. The strongest allele changed flowering time by approximately 3 days when homozygous under long day lengths. In both the maize and maize-teosinte populations, the QTL explaining the most variation in photoperiod response and expressing the strongest allelic effects mapped to chromosome 10. Genome-wide association and targeted high-resolution genetic mapping identified this QTL as a single gene, *ZmCCT*, a homologue of the rice photoperiod response regulator *Ghd7*. Under long day lengths *ZmCCT* alleles from diverse teosintes were consistently expressed at higher levels (2.9–12.6 times greater) and conferred later flowering than temperate maize alleles. *ZmCCT* therefore appears to act to repress expression of the maize florigen required to initiate flowering as *Ghd7* does in rice. Thus, the spread of maize to higher latitudes appeared to require selection after domestication by indigenous farmers of the Americas of rare mutations at *ZmCCT*, and a few other critical genes that reduced sensitivity to long day lengths. Subsequently Yang et al. (2013b) showed that insertion of a CACTA-like transposon into the *ZmCCT* promoter (2543 bp upstream of *ZmCCT*) suppresses *ZmCCT* expression, which in turn up-regulates the expression of the floral activator *ZCN8*, causing maize to flower early; in other words attenuates maize sensitivity under long day conditions. The authors used genome-wide association studies of maize and teosinte lines, sequencing of relevant regions of the *ZmCCT* gene and genetic transformation for proof of function. The transposable element (TE) insertion event occurred in a tropical maize plant and has been selected for and accumulated as maize adapted to long-day environments. As a result of this ‘selective sweep’ there is very low nucleotide diversity found around the insertion. This is a good example of a transposable element playing a key role in adaptive plant evolution and phenotypic variation by altering gene expression and function.

### **Maturity: Potato**

The potato was introduced from South America to Europe during the latter part of the sixteenth century, and from there to the rest of the world from the seventeenth century onwards. First in South America and then worldwide, locally adapted potatoes were selected that could produce tubers in different day lengths so that today potatoes are grown from latitudes 65°N to 50°S. Kloosterman et al. (2013) have shown that maturity in potato is controlled by the *StCDF1* locus on chromosome 5, and that selection of genetic variant *StCDF1.2* at this locus results in early

maturity and the ability to tuber in long days. They used fine mapping to locate *StCDF1* as a possible candidate gene and then used gene insertion by genetic transformation for proof of function. The early maturing/tuberizing *StCDF1.2* allele, under control of the cauliflower mosaic virus 35S promoter (CaMV 35S), was introduced into a short-day-dependent *S. tuberosum* group *andigena* genotype and into a very late maturing diploid potato genotype. Under long-day conditions, transformed Andigena plants had formed tubers by 4 weeks after planting, whereas the untransformed controls were unable to tuberize in long days. Overexpression of the *StCDF1.2* allele also resulted in early tuber formation in the late maturing diploid potato. Overexpression of *StCDF1.2* had no effect on flowering in potato, indicating a separation of signal transduction pathways for tuberization and flowering. Kloosterman et al. (2013) found that *StCDF1* belongs to the family of DOF (DNA-binding with one finger) transcription factors and regulates tuberization and plant life cycle length by acting as a mediator between the circadian clock and a mobile tuberization signal (StSP6A). In *StCDF1.2* an insertion results in a frame-shift which introduces a premature stop codon and results in a truncated protein with impaired ability to interact with other proteins, thus affecting post-translational regulation. The outcome is expression of the mobile StSP6A signal and induction of tuber development at the stolon termini in long days as well as short ones.

### **Vernalization: Wheat and Barley**

Winter wheats differ from spring wheats in their requirement for a long period at low temperatures to become competent to flower. This vernalization process prevents damage to the cold-sensitive flowering meristem during the winter. *VRN1* and *VRN2* are the main genes involved in the vernalization response in diploid wheat *Triticum monococcum* (Yan et al. 2003). However, in the economically important polyploid species of wheat, most of the variation in vernalization requirement is controlled by the *VRN1* locus. This gene is critical in polyploid wheats for their adaptation to autumn sowing and divides wheat cultivars into winter and spring types. The *VRN1* gene has been mapped in colinear regions of the long arm of chromosomes 5A, 5B and 5D. Yan et al. (2003) positionally cloned *VRN1* from diploid wheat and provided evidence from expression patterns for a completely linked MADS-box gene (*API*) being a strong candidate for *VRN1*. In *Arabidopsis*, *API* is expressed in apices and is one of the genes required for the transition between the vegetative and reproductive phases of growth. This transition is greatly accelerated by vernalization in wheat plants carrying the *vrn1* allele for winter growth habit. No *API* transcripts were detected in apices from unvernalized plants of *T. monococcum* with strong winter growth habit, but they were detected after 6 weeks of vernalization. In *T. monococcum* accessions with spring growth habit, *API* transcripts were observed in the apices without previous vernalization. Yan et al. (2003) explain their results with the following model (Table 5.2).

**Table 5.2** Genotypes and phenotypes for vernalization requirements in wheat: vernalization process reduces the abundance of the *Vrn2* gene product so that *vrn1* transcription gradually increases, leading to the competence to flower

<i>vrn1vrn1vrn2vrn2</i>	Spring habit	<i>vrn2vrn2</i> defective repressor
<i>Vrn1----vrn2vrn2</i>	Spring habit	<i>vrn2vrn2</i> defective repressor
<i>Vrn1----Vrn2----</i>	Spring habit	Repressor not recognized by <i>Vrn1</i> promoter
<i>vrn1vrn1Vrn2----</i>	Winter habit	<i>Vrn2</i> dominant repressor of <i>vrn1vrn1</i>

Gene product of *Vrn1* is same as *vrn1* and is required for flowering

The gene product of *Vrn1* is the same as *vrn1* and is required for flowering. *Vrn2* produces a repressor of *vrn1* and hence of flowering. Vernalization reduces the abundance of the *Vrn2* gene product so that *vrn1* transcription gradually increases, leading to the competence to flower. No vernalization is required if the repressor is defective/absent (*vrn2*) or not recognized by the promoter of the other gene (*Vrn1*). Gene sequencing confirmed that the predicted proteins from *vrn1* and *Vrn1* were either identical or differed by a single amino acid. In contrast, the promoter of *Vrn1* differed from that in three *vrn1* alleles by a 1-bp insertion and a 20-bp deletion located 728 and 176 bp, respectively, upstream from the start codon.

Yan et al. (2003) point out that a vernalization gene with a dominant spring growth habit has been mapped in the same map location in diploid wheat, barley and rye. Most of the wild Triticeae have a winter growth habit, suggesting that the recessive *vrn1* allele is the ancestral character. According to the model of Yan et al. (2003), independent mutations in the promoter regions of winter wheat, barley, and rye genotypes have resulted in the loss of the recognition site of the *VRN2* repressor (or an intermediate gene) and therefore in a dominant spring growth habit (*Vrn1* allele). Because this is a loss rather than a gain of a new function it is easier to explain its recurrent occurrence in the different Triticeae lineages.

Subsequently Yan et al. (2004) positionally cloned the wheat vernalization gene *VRN2* and confirmed its properties. Loss of function of *VRN2*, whether by natural mutations or deletions, resulted in spring lines, which do not require vernalization to flower. Reduction of the RNA level of *VRN2* by RNA interference accelerated the flowering time of transgenic winter-wheat plants by more than a month.

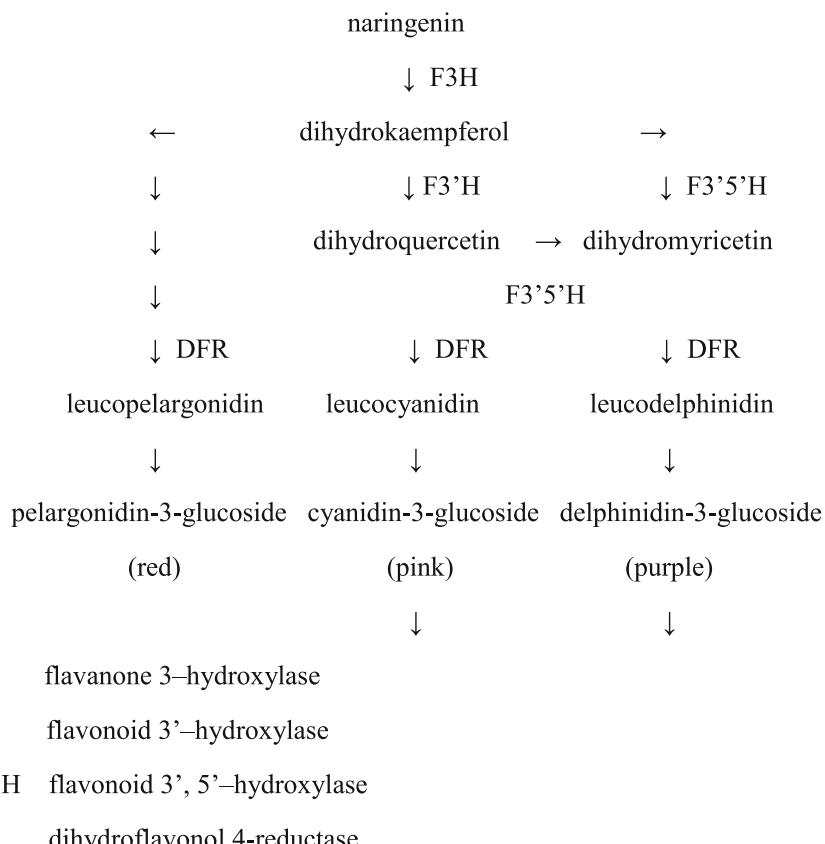
Later Yan et al. (2006) showed that a third ‘vernalization’ gene (*VRN3*) in wheat and barley is the *Arabidopsis FLOWERING LOCUS T (FT)*. The dominant allele *Vrn3* results in early flowering. In both wheat and barley, interactions occur for flowering time between vernalization treatment and *FT* allelic classes. Thus *FT* genes are responsible for natural allelic variation in vernalization requirement, providing additional sources of adaptive diversity to these economically important crops. Without going into details, Yan et al. (2006) proposed that *VRN3* interacts with *VRN2* in a similar manner to *VRN1*. In the absence of a functional *VRN2* repressor, the different mutations in the *VRN1* or *VRN3* regulatory regions have no effect on flowering. Furthermore, a mutation in a regulatory region of *VRN1* or *VRN3* is sufficient to preclude its recognition by the *VRN2*-mediated repression and to initiate the flowering cascade.

## Selection for Colour and Shape

### Colour: Tuber Skin of Potato

As early as 1910, Salaman (1910) identified three loci, *D* (developer), *R* (red), and *P* (purple) that determine the skin colour of potato tubers. One hundred years later, and Walter De Jong and his colleagues at Cornell University, USA, have confirmed the functions of these three loci. The *P* locus on chromosome 11 codes for flavonoid 3', 5'-hydroxylase (Jung et al. 2005); the *R* locus (*drf*) on chromosome 2 codes for dihydroflavonol 4-reductase (Zhang et al. 2009b); and the *D* locus on chromosome 10 encodes an R2R3 MYB transcription factor that regulates expression of multiple anthocyanin structural genes in tuber skin (Jung et al. 2009). These papers also review and discuss previous genetic and chemical analysis of pigment colour in the flowers, tuber skin and tuber flesh of both diploid and tetraploid potatoes. Here, however, a very brief summary for tetraploids is in order. Red and purple tuber anthocyanins are derivatives of pelargonidin and delphinidin, respectively. *R* is required for the production of red anthocyanins, *P* is required for the production of purple anthocyanins, and *D* is required for the tissue-specific accumulation of anthocyanin pigment in tuber skin. Tubers with genotype *D*--- *R* --- *P*--- or *D*--- *rrrr* *P*--- have purple skin, those with genotype *D*--- *R* --- *pppp* have red skin, and those with genotype *dddd* ---- ---- are white. As purple-skinned potatoes can have genotype *R*--- or *rrrr*, both the 'red' and 'not-red' alleles of *drf* must code for catalytically-active enzymes. The enzyme encoded by the red allele appears capable of reducing dihydromyricetin as well as dihydrokaempferol, whereas the non-red allele(s) appears capable of reducing the former, but not the latter (Fig. 5.2). In contrast, the *pppp* genotype is a loss of function of enzyme activity. The R2R3 MYB domain gene is transcribed in red and purple, but not in white, tuber skin. Jung et al. (2009) think that there is now good evidence that evolution and subsequent selection for novel tissue-specific expression patterns of R2R3 MYB domain genes is responsible for much of the observed anthocyanin-mediated variation in the *Solanaceae*. As mentioned earlier, Mendel's flower colour gene (*A*) is also a (bHLH) transcription factor that controls the anthocyanin biosynthesis pathway.

A key feature of the potato research done by the group at Cornell was direct evidence of gene function, by showing that the relevant dominant allele could transgenically complement the recessive allele in respect of both tuber colour and anthocyanin composition. In other words, genetic transformation was used to introduce the dominant allele into a potato with the recessive allele. In these examples, the potato genes were cloned and introduced back into potato by genetic transformation. In Chap. 17 we shall look at the introduction of novel biochemistry.



**Fig. 5.2** Part of anthocyanin biosynthetic pathway, simplified (Holton and Cornish 1995)

## *Colour: Red and White Grapes*

Grapevines (*Vitis vinifera* L.) have been cultivated for thousands of years for fresh fruit, dried fruit and wine production (Walker et al. 2007). The progenitors of modern grapevines are considered to have had red/black (red) berries, but by Roman times white grapes were cultivated using clonal cuttings as described by Pliny the Elder in 77 CE (AD). Today over 7000 clonally propagated cultivars are recognized, about half of which have white berries. The colour difference between red and white grapes is based on the presence or absence of anthocyanin in the berry skin, and appeared to be controlled by a single locus, with white the recessive trait. The *VvUFGT* (*Vitis vinifera* UDP-glucose 3-*O*-flavonoid:glucosyltransferase) gene, which encodes an enzyme catalysing one of the last steps of the anthocyanin biosynthetic pathway, is expressed at much lower levels in white grapes than in red fruit. However, no significant differences were detected between the *VvUFGT* gene sequence from red and white cultivars, suggesting that a difference in regulation of

*VvUFGT* transcription is responsible for the different berry colours. Walker et al. (2007) found that the berry colour locus comprises two very similar tightly linked regulatory genes, *VvMYBA1* and *VvMYBA2*, both of which need to be inactivated to prevent anthocyanin biosynthesis. *VvMYBA1* is inactivated in white berries by the presence of a retrotransposon in its promoter (Kobayashi et al. 2004). In contrast, the white berry allele of *VvMYBA2* contains two non-conservative mutations, one leading to an amino acid substitution and the other to a frame shift resulting in a smaller protein. A transient assay method was developed to determine the role of each of the *VvMYBA* genes, in which the function of the genes could be assessed using the activation of anthocyanin biosynthesis and two reporter systems. Expression constructs containing either 35S:*VvMYBA1* or 35S:*VvMYBA2* (*red alleles*) were capable of activating anthocyanin biosynthesis and *GFP* (*Green Fluorescent Protein*) expression to the same extent, suggesting that both coding sequences are capable of acting in a similar fashion.

Walker et al. (2007) concluded from an extensive analysis of 55 white and 60 red cultivars that all extant white cultivars of grapevines have a common origin because the same mutations were present in all the white cultivars examined. Hence rare mutational events in two adjacent genes were essential for the genesis of the white grapes used to produce the white wines and white table grapes we enjoy today. After duplication of the *VvMYBA* genes, *VvMYBA2r* and *VvMYBA1r* were both active resulting in a red grape phenotype. The accumulation of two mutations in *VvMYBA2w* and the insertion of a retrotransposon in the promoter of *VvMYBA1w* results in the inactivation of both these genes. However, the phenotype of the berries would still be red (dominant trait) because the ancestral plant would be heterozygous for the red and white alleles. Segregation of the red and white alleles during sexual reproduction resulted in homozygous white-fruited seedlings, the antecedents of modern white cultivars.

### ***Shape: Tomatoes***

In Chap. 1 we saw that the cultivated tomato (*Solanum lycopersicum*) fruit is highly diverse in shape and colour as well as having increased weight compared with its wild relatives that carry round and two-loculed fruits. The fruits of the cherry tomato, tomato var. *cerasiforme*, are typically larger than those of wild species but smaller than those of cultivated tomato. Today tomatoes can be classified into eight fruit shape categories: flat, round, rectangular, ellipsoid, heart, long, ovoid and oxheart. About 70 % of the variation in fruit shape found in a diverse collection of 368 tomato accessions (7 wild, 46 *cerasiforme* and 315 cultivated tomatoes from Italy, Spain, Latin America and USA) could be explained by combinations of variants that arose in just four genes (Rodríguez et al. 2011). *SUN* and *OVATE* control elongated shape whereas *FASCIATED* (*FAS*) and *LOCULE NUMBER* (*LC*) control fruit locule number and flat shape. Rodríguez et al. (2011) summarized knowledge of these genes as follows. *SUN* encodes a protein that is a positive

regulator of growth resulting in elongated fruit and is hypothesized to alter hormone or secondary metabolite levels. The mutation is the result of a gene duplication event that was mediated by the retrotransposon *Rider*. *OVATE* encodes a negative regulator of growth, presumably by acting as a repressor of transcription and thereby reducing fruit length. The *OVATE* allele that conditions an elongated fruit carries a premature stop codon and is presumed to be a null allele. Locus number, which has a pleiotropic effect on fruit shape and size, is controlled by the fasciated (*fas*) and locule number (*lc*) loci. *FAS* encodes a YABBY transcription factor and down-regulation of the gene is caused by a large insertion in the first intron (estimated to be 6–8 kb), resulting in fruits with high locule number. Two single-nucleotide polymorphisms (SNPs) were found to be critical in controlling the locule number (LC) phenotype and were located approximately 1200-bp downstream of the stop codon of a gene encoding a WUSCHEL homeodomain protein, members of which regulate stem cell fate in plants. All obovoid, and many of the ellipsoid (83 %), rectangular (59 %), and heart (48 %) tomatoes carried the mutant allele of *OVATE* whereas most of the long (88 %) and oxheart (83 %) tomatoes carried the mutant allele of *SUN*. The most frequent mutation in flat tomatoes was *LC* (82 %) followed by *FAS* (28 %). Many of the long tomatoes also carried the mutation in the *LC* gene (63 %). All oxheart tomatoes carried the *LC* mutation in addition to *SUN* and/or *FAS*. Most round tomatoes carried the wild-type allele at the four shape loci, with the *LC* mutation most prevalent at 33 %. The majority of the round tomatoes with the *LC* mutation were tomato var. *cerasiforme* lines. Rodríguez et al. (2011) concluded that the *LC*, *FAS*, and *SUN* mutations arose in the same ancestral population while the *OVATE* mutation arose in a separate lineage. Furthermore, *LC*, *OVATE* and *FAS* mutations may have arisen prior to domestication or early during the selection of cultivated tomato, whereas the *SUN* mutation appeared to be a post domestication event arising in Europe. The first written record of tomato in Europe was in 1544 where it was described as having flat and segmented fruit. Other descriptions of fasciated fruit soon followed. The fascinated phenotype of those tomatoes suggests that the earliest tomatoes to arrive in Europe carried both *LC* and *FAS* mutations.

## Green Revolution Genes

### *Wheat*

World wheat grain yields increased substantially in the 1960s and 1970s because farmers rapidly adopted the new cultivars and cultivation methods of the so-called ‘green revolution’. The new cultivars were shorter, increased grain yield at the expense of straw biomass and were more resistant to damage by wind and rain (McVittie et al. 1978; Peng et al. 1999). These wheats are short because they respond abnormally to the plant growth hormone gibberellin. This reduced response

to gibberellin is conferred by semidominant, mutant dwarfing alleles at one of two 'Reduced height-1' (*Rht-B1* and *Rht-D1*) loci, on chromosomes 4B and 4D, respectively. They are altered function rather than loss-of-function mutant alleles which reduce plant height, reduce responses to gibberellins and increase *in planta* gibberellins levels. The two most widely used dwarfing alleles are *Rht-B1b* and *Rht-D1b*, both derived from Norin 10. Another source of extreme dwarfism is cultivar Tom Thumb, but it has been less widely used in the production of commercial dwarfs (McVittie et al. 1978). Peng et al. (1999) showed that *Rht-B1/Rht-D1* and maize dwarf-8 (*d8*) are orthologues (copies because have same evolutionary origin) of the 'Arabidopsis Gibberellin Insensitive' (*GAI*) gene. These genes encode proteins that resemble nuclear transcription factors and contain an SH2-like10 domain, indicating that phosphotyrosine may participate in gibberellin signalling. Six different orthologous dwarfing mutant alleles encode proteins that are altered in a conserved amino-terminal gibberellin signalling domain. The *gai* allele encodes a mutant protein (*gai*), lacking 17 amino acids from near the amino terminus, that is thought to confer the altered gibberellin responses characteristic of the *gai* mutant. Transgenic rice plants containing a mutant *GAI* allele give reduced responses to gibberellin and are dwarfed, indicating that these dominant mutant *GAI* orthologues could be used to increase yield in a wide range of crop species. Wheat breeders need to determine the best combinations of dwarfing genes/alleles in particular genetic backgrounds for desired height and high yield.

## Rice

The recessive semidwarfing gene *sd-1* gene was first identified in the Chinese variety Dee-geo-woo-gen (DGWG), which was crossed in the early 1960s with Peta (tall) to develop the semidwarf cultivar IR8. This cultivar produced record yields throughout Asia and formed the basis for the development of new high-yielding, semidwarf plant types. Since the 1960s, *sd-1* has remained the predominant semidwarfing gene in rice cultivars. Its presence results in a shortened culm with improved lodging resistance and a greater harvest index, allowing increased use of nitrogen fertilizers. Its phenotype is consistent with the dwarfism that results from a deficiency in bioactive GA1 and *sd-1* semidwarf rice retains the ability to respond to applications of bioactive gibberellins. Spielmeyer et al. (2002) provided convincing evidence that the gibberellin (GA) 20-oxidase gene (*Os20ox2*) corresponds to the *Sd-1* locus. Sequence data from the rice genome was combined with previous mapping studies to locate a putative GA 20-oxidase gene (*Os20ox2*) at the predicted map location of *sd-1* on chromosome 1. Two independent *sd-1* alleles contained alterations within *Os20ox2*. A deletion of 280 bp within the coding region of *Os20ox2* was predicted to encode a non-functional protein in an *indica* type semidwarf (Doongara). A substitution in an amino acid residue (Leu-266) that is highly conserved among dioxygenases could explain loss of function of *Os20ox2* in a *japonica* semidwarf (Calrose76). Quantification of GAs in elongating stems by

GC-MS showed that the initial substrate of GA 20-oxidase activity (GA53) accumulated, whereas the content of the major product (GA20) and of bioactive GA1 was lower in semidwarf compared with tall lines. Asano et al. (2007) found six different *sd-1* alleles in 38 (15 were IR8 allele) out of 57 semi-dwarf cultivars of rice, both *japonica* and *indica*, from China, USA and Japan.

## Genome-Wide Analysis of Gene Expression

In the examples considered so far in this chapter, we have looked at the expression of relatively few genes (alleles) with clear and desirable phenotypes that were selected by farmers and plant breeders. In recent years new technologies have developed for genome-wide analysis of gene expression patterns. Transcriptomics is a highly automated process that allows experimenters to identify all of the changes in gene expression that occur in response to a particular treatment or at a particular stage of plant development. This may help to identify genes that are potentially useful to plant breeders and biotechnologists. For example, genes that are expressed in high amounts following an attack by a pest or pathogen may be important in the plant's defence. Furthermore, comparisons in gene expression can be made between germplasm with desirable and undesirable attributes in order to seek desirable alleles. In proteomics, the next level of gene expression is studied, namely the total protein complement of cells and tissues. It is heavily dependent on electrophoretic separation of proteins in either one or two dimensions. Finally metabolomics allows the study of the chemical constituents (metabolites) of cells using the methods of analytical chemistry, particularly gas chromatography (GC) coupled to mass spectrometry (MS). Different classes of metabolites can be studied in a range of germplasm and variation explained in terms of biochemical pathways and candidate genes, and correlated with achieving desirable (physiological) traits. One interesting application is with genetically modified (GM) crops. Differences between a genetically modified and non modified cultivar can be compared with differences between cultivars of the same species in order to quantify the effects of genetic modification on chemical composition and hence food safety. The interested reader can find detailed information on these 'omics' technologies in textbooks on plant biotechnology. Here we will finish with a brief consideration of microarrays.

### ***Microarrays***

Microarrays are a tool for monitoring the gene expression levels of thousands of genes in parallel. This following brief account is taken from *Microarray Gene Expression Data Analysis* by Causton et al. (2003). A microarray is typically a glass or polymer slide onto which DNA molecules are attached at fixed locations called

spots. There may be tens of thousands of spots on an array, each containing tens of millions of identical DNA molecules or fragments of identical molecules, of lengths from tens to hundreds of nucleotides. For gene expression studies, each of these molecules should identify a single mRNA molecule, or transcript, in a genome. The spots are either printed on the microarrays by a robot or jet, or synthesized *in situ* by photolithography or by inkjet printing. One common microarray application is to compare the gene expression levels in two different samples of interest, e.g. the same cells or cell type under two different conditions, or two different genotypes. A representation of the mRNA extracted from each of the samples is labelled in two different ways; for example, a green dye for one sample and a red one for the other sample. This can be done using a reverse transcriptase enzyme (RT) that generates a complementary cDNA to the mRNA. During that process fluorescent nucleotides are attached to the cDNA. The technology is based on the binding (hybridization) of two complementary single-stranded nucleic acid molecules, namely cDNA and the complementary encoding strand of the gene (DNA); or mRNA to the template strand. Two single-stranded nucleic acid molecules that are not fully complementary may also hybridize, but the greater the complementarity, the stronger the binding. The hybridized microarray is excited by a laser and scanned at wavelengths suitable for the detection of the green and red dyes. The amount of fluorescence emitted upon laser excitation corresponds to the amount of cDNA (mRNA) bound to each spot. If the cDNA (mRNA) from sample one is abundant the spot will be green whereas it will be red if sample two is in abundance. If both are equal the spot will be yellow and if neither is present it will not fluoresce and appear black. Hence the fluorescence intensities and colours for each spot reveal the relative expression levels of the genes in the two samples. Information processing then takes place to find differentially expressed genes or clusters of similarly expressed genes and to formulate new hypotheses about underlying biological processes. In the context of plant breeding, the aim is to discover genes of economic importance and then seek desirable variants for crop improvement.

## Heritable Epigenetic Variation Among Maize Inbreds

I am going to finish this chapter on gene expression with an example of a phenomenon that is attracting a lot of research attention in many organisms, namely heritable epigenetic variation. Epigenetic variation describes heritable differences in gene expression that are not attributable to changes in DNA sequence. Such variation can occur in the absence of any genetic change or in addition to genetic differences. Methylation of cytosine residues provides one mechanism for the inheritance of epigenetic information. It often acts to suppress the activity of transposable elements, repetitive sequences, pseudogenes, and in some cases otherwise active genes; and it may play a role in generating variation that could provide adaptation to environmental stresses (Eichten et al. 2011). However, the abundance and role of epigenetic variation has not yet been well characterized.

Eichten et al. (2011) used genome-wide profiling of DNA methylation patterns to assess the relationship of methylation to chromosomal and gene structures in two maize inbred lines, B73 and Mo17. The details of the method need not concern us, but involved an array platform of oligonucleotide probes and methylated DNA immunoprecipitation (meDIP) on fragmented genomic DNA using a 5-methylcytosine antibody. Although the majority of the genome showed highly similar methylation patterns in both inbreds there were several hundred differentially methylated regions (DMRs) found throughout the maize genome. In general, many of these DMRs were located in intergenic regions and may reflect differences in transposon silencing among the genotypes. However, at least 71 of the 690 variable methylation regions were found within 500 bp of a high-confidence gene (FGS). These regions were characterized further using a population of near isogenic lines (NILs). In general, relatively stable inheritance of DMRs was found with rare examples of both gains and losses of DNA methylation. Thirteen DMRs were studied in detail: three of the regions exhibited evidence for *trans*-acting control of DNA methylation patterns; the remaining ten loci had methylation patterns that were either stably inherited or reflected *cis*-linked genetic changes that were directing the methylation difference. The analyses of several identical-by-descent regions (shared inheritance of same DNA sequence from common ancestor) of the B73 and Mo17 genomes also provided evidence that epigenetic variation can occur in the absence of nearby genetic polymorphisms. Overall, the authors have provided evidence for faithfully inherited methylation differences that would not be revealed by high-resolution analyses of genetic differences. The functional consequences and contribution to phenotypic variation are still to be determined; plant breeders await the results and implications for their work with interest.

# **Chapter 6**

## **Quantitative Genetics and Genomic Selection**

### **Introduction**

Most economically important traits display continuous variation and hence individual genes affecting quantitative traits cannot be recognized by Mendelian analysis because discrete classes cannot be identified. Nevertheless, Nilsson-Ehle (1909) and East (1915) independently realized that the segregation of a number of Mendelian factors (genes) of similar and cumulative action, together with the effects of non-heritable factors, could account for continuous variation. Furthermore, Fisher (1918) showed that genetic information could be inferred from measurements on related individuals using biometrical models which assumed Mendelian inheritance. Subsequently this biometrical approach was refined to take account of all the properties of genes known from Mendelian genetics, such as dominance, epistasis and linkage. In later parts of this book, when necessary, we will incorporate these properties into our models for quantitative traits. Methods of analysis can be found in books such as *The Genetical Analysis of Quantitative Traits* (Kearsey and Pooni 1996), the fourth edition of *Introduction to Quantitative Genetics* (Falconer and Mackay 1996) and *Genetics and Analysis of Quantitative Traits* (Lynch and Walsh 1998). Quantitative Genetics has provided the foundation for efficient breeding methods which are still the main route to new cultivars. Today genes affecting quantitative traits can be detected indirectly through associations between trait scores and molecular markers with known positions on chromosomes, and are referred to as Quantitative Trait Loci (QTLs). Before continuing, the reader may find Box 6.1 a useful reminder of some statistics that are used throughout the rest of this book. As we will primarily be concerned with theoretical considerations, whole population statistics are given rather than sample estimates. Statisticians distinguish between random and fixed variables, the latter being determined without sampling error by the experimenter. In practice, if we want to predict variable 'y' on the basis of knowledge about variable 'x', we talk about the regression of y (random) on x (fixed), and assume a linear-regression model. If on the other hand

we are interested in the degree of relationship between two variables ‘ $x$ ’ and ‘ $y$ ’, we speak of the correlation between  $x$  (random) and  $y$  (random) and assume a bivariate-normal model. Before the advent of modern computers, every effort had to be made in the design of genetic and breeding experiments to accommodate the requirements of conventional statistical techniques such as the analysis of variance (ANOVA) and multiple regression analysis, for example treatments with equal sample sizes and no (or very little) missing data. Analysis of variance and multiple linear regression analysis are in fact both special cases of the General Mixed Model in which unequal sample sizes and missing data can be handled in a relatively straightforward way by suitable software on fast and powerful computers.

### Box 6.1

Statistics used in this and subsequent chapters, where  $x$  and  $y$  are two measured variables and  $y$  is a random variable whereas  $x$  may be a random or fixed variable

Sum of  $n$  values :  $x_1 + x_2 + x_3 + \dots + x_n = \sum x$

Population mean :  $\mu_x = (\sum x)/n;$

$= \sum f x$  when  $f$  is proportion of  $x$  and hence  $\sum f = 1$

Variance  $V_x = \sigma_x^2 = \left( \sum (x - \mu_x)^2 \right)/n;$

$= \sum f(x - \mu_x)^2 = \sum f x^2 - \mu_x^2$ , when  $f$  is proportion of  $x$

$V\mu_x = V_x/n$

Covariance  $Cov_{xy} = \left( \sum (x - \mu_x)(y - \mu_y) \right)/n;$

$= \sum f(x - \mu_x)(y - \mu_y) = \sum f xy - \mu_x \mu_y$ ,  $f$  is proportion

#### Regression of $y$ (random) on $x$ (fixed): linear regression model

Linear regression model:

$$\hat{y} = b_0 + b_1 x$$

where  $\hat{y}$  is the predicted value of  $y$ ,  $b_1$  is the slope of the regression line and  $b_0$  is the intercept, with  $b_0$  and  $b_1$  chosen (estimated) to minimize  $\sum (y - \hat{y})^2$  (least squares estimate).

$$\begin{aligned} b_0 &= \mu_y - b_1 \mu_x \\ b_1 &= Cov_{xy}/V_x \end{aligned}$$

(continued)

**Box 6.1** (continued)Correlation of  $x$  (random) and  $y$  (random): bivariate normal model

$$r = \text{Cov}_{xy}/(\sigma_x\sigma_y)$$

measures the degree to which the actual values of  $y$  agree with the predicted values  $\hat{y}$ , with range  $-1$  (perfect negative relationship) to  $+1$  (perfect positive relationship), where  $0$  is no relationship.

$$r^2 = \left( \sum (\hat{y} - \mu_y)^2 \right) / \left( \sum (y - \mu_y)^2 \right)$$

i.e. the proportion of the sum of squares for ‘ $y$ ’ accounted for by the regression on ‘ $x$ ’

Multiple linear regression ( $p$  fixed or random variables, latter multivariate-normal)

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + \dots + b_px_p$$

$b_0$  intercept,  $b_1, b_2 \dots b_p$  are the regression coefficients for predictors  $x_1, x_2 \dots x_p$ , respectively, and are estimated to minimize  $\sum (y - \hat{y})^2$  (least squares estimate).

If all of the variables are standardized to a mean of zero and standard deviation of one, the intercept is equal to zero and the regression coefficients are called standardized regression coefficients ( $\beta$ ).

The multiple correlation coefficient  $R$  is the correlation between  $y$  and  $\hat{y}$  ( $R$  has values between  $0$  and  $1$ ).  $R^2$  is the percentage of variation accounted for by the  $x$ ’s, but the sample estimate is not an unbiased estimate of the population parameter (needs correction for  $p$  and  $n$ ).

## Quantitative Traits

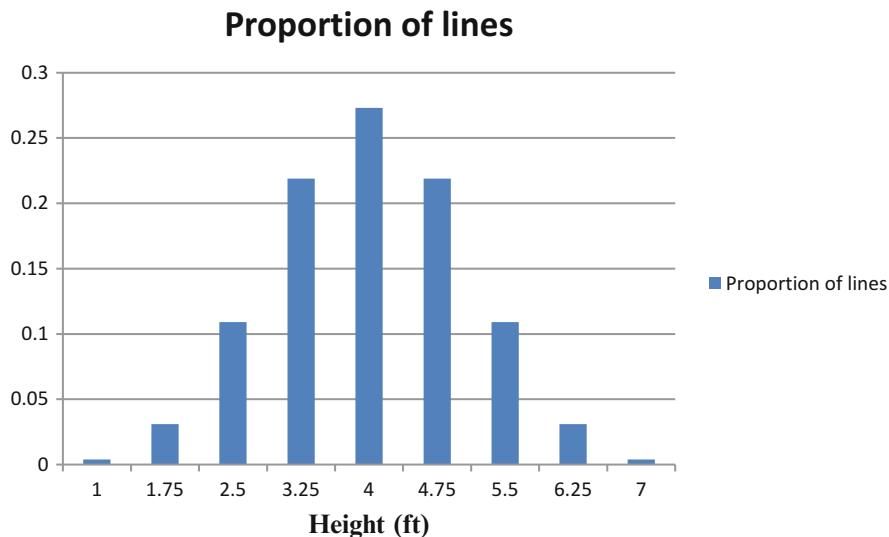
Height was one of the traits that Mendel analysed in his garden peas, as dealt with in Chaps. 4 and 5. Mendel crossed a true-breeding tall variety (6–7 ft tall) with a true-breeding short one ( $\frac{3}{4}$  to  $1\frac{1}{2}$  ft) and found that the resulting hybrids were tall (6 to  $7\frac{1}{2}$  ft). On self-pollination, Mendel observed 787 tall and 277 short plants in the next generation, which was not significantly different from a 3:1 ratio ( $X^2 = 0.61$ ;  $P > 0.05$ ). The differences in height were so large that Mendel could categorize his plants as tall and short. Nevertheless, one could also say that the difference between genotypes  $AA$  and  $aa$  was about 6 ft, and hence this was the effect of replacing two copies of ‘ $a$ ’ with two copies of ‘ $A$ ’. Furthermore, as  $AA$  and  $Aa$  were about the same height (complete dominance), just one copy of ‘ $A$ ’ was sufficient to increase height by about 6 ft. One could predict that continuous self-pollination over

**Table 6.1** Proportions of true breeding lines for eight unlinked genes of equal effect on height

Increasing allele	Decreasing allele	Probability	$P = Q = \frac{1}{2}$	Height ft	Genotype
0	8	$Q^8$	0.00390625	1.00	<i>aabbccddeeffgghh</i>
1	7	$8PQ^7$	0.03125	1.75	<i>AAbbccddeeffgghh</i> etc.
2	6	$28P^2Q^6$	0.109375	2.50	<i>AABBccddeeffgghh</i> etc.
3	5	$56P^3Q^5$	0.21875	3.25	<i>AABBCCcddeeffgghh</i> etc.
4	4	$70P^4Q^4$	0.2734375	4.00	<i>AABBCCDDeeffgghh</i> etc.
5	3	$56P^5Q^3$	0.21875	4.75	<i>AABBCCDDEEfFgghh</i> etc.
6	2	$28P^6Q^2$	0.109375	5.50	<i>AABBCCDDEEFFgghh</i> etc.
7	1	$8P^7Q$	0.03125	6.25	<i>AABBCCDDEEFFGGhh</i> etc.
8	0	$P^8$	0.00390625	7.00	<i>AABBCCDDEEFFGGHH</i>

generations would eventually produce equal numbers of true-breeding tall (*AA*) and short (*aa*) lines. Genes of large effect are known to affect height in a number of species, such as the dwarfing genes in cereals mentioned in Chaps. 2 and 5 under The Green Revolution. But height can display continuous variation, as usually found with yield and many quality traits.

So now let us suppose that eight unlinked genes of equal effect explained the difference between the tall and short variety (*AABBCCDDEEFFGGHH*  $\times$  *aabbccddeeffgghh*) so that the difference between *AA* and *aa* is  $\frac{3}{4}$  ft, where 'A' is the allele of increasing effect, and likewise for all of the other genetic differences. The set of true breeding lines that can be obtained from the cross forms a binomial distribution in terms of the number of alleles of increasing effect, as shown in Table 6.1 and Fig. 6.1. With environmental as well as genetic variation, the nine discrete classes would merge into a continuous distribution of heights, and approximate to a bell-shaped normal distribution (Fig. 6.1). Incidentally, the normal distribution was discovered in 1733 by Abraham de Moivre as an approximation to the binomial distribution when the number of trials is large, although the distribution is more usually associated with the name of Carl Friedrich Gauss who derived it in 1809 (Bulmer 1967). If  $X$  is normally distributed with mean  $\mu$  and variance  $\sigma^2$ , then  $Z = (X - \mu)/\sigma$  follows a normal distribution with zero mean and unit variance.  $Z$  is called a standard normal variate whose density function and cumulative probability function respectively can be found in Statistical Tables (Fisher and Yates 1963). In our example we have a binomial distribution with mean 4 and variance 1.125. We can compare the binomial proportions with those determined for a normal distribution with the same mean and variance for the intervals 1.375–2.125, 2.125–2.875 ... 5.875–6.625 (i.e. the areas under the normal curve between these pairs of ordinates, in standard deviation units from the mean, determined from the cumulative probability function of the standard normal distribution). It can be seen in Table 6.2 that there is good agreement between the exact binomial (a discrete function) results and those from the normal (a continuous



**Fig. 6.1** Proportions of true breeding lines for eight unlinked genes of equal effect on height

**Table 6.2** Proportions of true breeding lines for eight unlinked genes of equal effect on height

Height ft	Binomial proportion	Height ft	Normal proportion
1.00	0.00391	<1.375	0.00666
1.75	0.03125	1.375–2.125	0.03170
2.50	0.10938	2.125–2.875	0.10621
3.25	0.21875	2.875–3.625	0.21673
4.00	0.27344	3.625–4.375	0.27741
4.75	0.21875	4.375–5.125	0.21673
5.50	0.10938	5.125–5.875	0.10621
6.25	0.03125	5.875–6.625	0.03170
7.00	0.00391	>6.625	0.00666

function) approximation. The normal distribution is useful in quantitative genetics because the sum of two independent normal variates is itself normally distributed (Bulmer 1967). Furthermore, it can be shown that the sum of a large number of independent random variables are approximately normally distributed almost regardless of their individual distributions, a result known as the central limit theorem (Bulmer 1967). For example, we have assumed that height is determined by a large number of factors, both genetic and environmental, which are additive in their effects. If they were multiplicative, then the logarithms of the effects would be additive and we should expect the logarithm of the end product to be normally distributed. Hence an appropriate scale of measurement can be an issue. Now, however, we need to look at the fact that quantitative traits display environmental variation as well as genetic variation.

## **Yield Trials**

In modern farming practice crops are machine planted or sown in rows, usually in relatively large fields of a few hectares. Also the crop is usually a monoculture of a single cultivar, which is genetically uniform: for example, a clone of a potato genotype, an inbred line of wheat or barley, or an  $F_1$  hybrid of maize. Despite modern methods of precision and controlled traffic farming, the field will be heterogeneous in terms of soil fertility and moisture. If at harvest we took the trouble to separately harvest and weigh each row in short sections, we would find differences in yield due to soil heterogeneity, although commonly sections in close proximity would be more alike than those further apart. In other words, quantitative traits such as yield display environmental variation. Differences will also be seen when the cultivar is grown in different fields and in more than one location or year. The same will be true of protected crops grown under glass; the environment may be controlled, but it will not be completely uniform.

How then can we be sure that new cultivars are really better than existing ones, and that the potential cultivars we select as breeders are the best of those available? Potential cultivars (genotypes) are assessed in yield trials, the aim being to accurately predict their relative yields and other attributes when grown under agricultural or horticultural conditions. Interestingly, Simmonds (1981) provided evidence for two crops, potatoes and sugarcane, that trials are not good predictors of agricultural yields. Hence duplication of farmers' fields and management procedures are important for the relevance of the results. Let's start by considering a yield trial in a field at one location in 1 year. The trial area will be divided into a rectangular grid of plots in which to grow the entries (Fig. 6.2). More than one plot of each entry is required to assess the environmental variation because the differences between entries in different plots will be due to environment ( $\sigma^2$ ) as well as genotype ( $\sigma^2_G$ ). Good experimental techniques are required to minimize the increase in 'environmental' variation from lack of uniformity in plant spacing and the application of fertilizers, herbicides, fungicides and pesticides. In the early generations of a breeding programme, and in trials for genetic analysis, there will be a large number of entries, often several hundred. Furthermore, plots will be small in size because only limited amounts of planting material will be available. At most, plots will comprise several rows, but sometimes a single row or even a single plant may be used. For a given total trial area, statistical considerations suggest that the use of many smaller plots is preferable to fewer larger ones (Mead 1997), but shape and size will often be dictated by agricultural practice and machinery. Furthermore, resources usually permit relatively few replicates of each entry, often two, three or four. The simplest experimental layout would be to allocate all of the entries and their replicates at random to the trial plots; that is complete individual randomization. Randomization is necessary for a valid estimate of the precision of a trial which is the standard error of the difference between the means (over replicates) of two entries; the smaller the standard error the better.



**Fig. 6.2** Replicated yield trials of potato at ware site in Dundee, Scotland (*source:* SCRI)

A slightly more complicated, but standard layout, is to have each replicate in a separate block so that each block contains a single plot of each entry (randomized complete block). A different randomization of entries is used in each block. This leads to a simple additive model of plot yield ( $y$ ):

$$y = \mu + b + g + e$$

where  $\mu$  is the trial mean,  $b$  is the block effect,  $g$  is the genotype effect and  $e$  is the plot error.

The assumption of additivity implies that both the differences between blocks and the differences between genotypes remain consistent over a trial. A second assumption is usually made in the analysis of trials, namely that plot errors are homogeneous in size and unaffected by genotype; that is the yield variation between plots is similar for all of the genotypes in the trial. The standard method of analysis is based on least squares estimates for general linear models (Mead 1997). Thus one can estimate from the mean squares in an analysis of variance (ANOVA) of plot yields, the plot to plot variation (soil heterogeneity) for replicates of the same genotype ( $\sigma^2$ ) and the variation between the mean (averaged over replicates) yields of genotypes ( $\sigma_G^2 + \sigma^2/r$ ), where  $r$  is the number of replicates. The ratio  $\sigma_G^2/(\sigma_G^2 + \sigma^2/r)$  is called the broad sense heritability ( $h_b^2$ ), and measures the proportion of the observed (phenotypic) variation between potential cultivar means that is due to genetic (genotype) differences. The standard error (SED) for

comparing two genotype means is  $[(2\sigma^2/r)]^{1/2}$ . Hence the broad sense heritability can be increased, and the standard error for comparing two genotypes can be decreased, by increasing the number of replicates.

Two approaches can be taken to reduce the problem of within-trial heterogeneity to reduce  $\sigma^2$ . One is to group plots into small blocks for which the assumption of uniform conditions within blocks is more realistic (Mead 1997). The other is the use of spatial analysis (nearest neighbour) which models the observed similarity of yields from adjacent plots (Gleeson 1997). Computer generated trial designs and software for their analysis are available for both approaches, for example AGROBASE Generation II™ from Agronomix Software (<http://www.agronomix.com>). Blocking is highly desirable and can be supplemented with spatial analysis, although the latter will not be considered further in this book.

Failure to use blocks will usually result in a mixture of two disadvantages. Firstly, the distortion of relative means, because by chance all replicates of some genotypes are on high fertility plots whereas for others all replicates are on low ones; hence the differences between genotype means will be biased. Secondly, the inflation of standard errors, because by chance the replicates of some genotypes are on an equal number of high and low fertility plots; hence the plot to plot variation is large for these genotypes. The effective control of variation through blocking usually requires fairly small block sizes, whilst most selection trials use large numbers of genotypes. This leads almost inevitably to the use of incomplete block designs. Thus a genotype trial will be designed in complete replicate blocks, each replicate being split into a number of incomplete blocks. Comparisons between genotype means will then require an adjustment to allow for differences between blocks, as explained in books on the subject (Mead et al. 2002). The subset of genotypes in each incomplete block in one replicate should be evenly spread among blocks in each other replicate. Alpha designs, also called generalized lattices (Patterson et al. 1978; Patterson et al. 1988), extended the lattice designs of Yates (1936) and have provided a well-understood system for producing appropriate and efficient designs; for example both in Recommended List cereal cultivar trials in the UK (Patterson and Hunter 1983) and in breeders' small plot barley trials in the UK (Robinson et al. 1988). In the former, using incomplete blocks rather than complete blocks reduced the variances of cultivar yield differences by an average of 30 %, which can be viewed as the gain in efficiency. Analysis of data from past genotype trials suggests that allowing for block differences in two directions can further improve precision. The lattice squares of Yates (1937) required the number of cultivars to be a perfect square  $p^2$  and the number of replicates to be  $p + 1$ , and hence were rather restrictive. Today other row-and-column designs are available which place fewer restrictions on the numbers of cultivars and replicates (Patterson and Robinson 1989). A  $10 \times 10$  design with two replicates is shown in Fig. 6.3, where it is easily seen that cultivars 1–10 appear in different rows in replicate 2 and share their columns with nine different cultivars.

## Replicate 1

1	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30
31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48	49	50
51	52	53	54	55	56	57	58	59	60
61	62	63	64	65	66	67	68	69	70
71	72	73	74	75	76	77	78	79	80
81	82	83	84	85	86	87	88	89	90
91	92	93	94	95	96	97	98	99	100

## Replicate 2

1	20	28	36	49	57	64	72	83	95
17	65	43	89	6	100	78	54	21	32
26	4	52	80	15	88	99	63	37	41
40	48	85	51	74	2	16	97	69	23
44	22	7	98	33	79	90	11	55	66
59	73	96	47	30	61	5	38	12	84
68	87	19	3	92	24	31	45	76	60
75	91	34	62	58	13	27	86	50	9
82	39	71	25	67	46	53	10	94	18
93	56	70	14	81	35	42	29	8	77

**Fig. 6.3** Two replicates, 10 rows by 10 columns design for field trial comparing 100 cultivars (numbered 1–100) in small plots (Patterson and Robinson 1989)

## The General Mixed Model and Estimation of Breeding Values

Comparisons of breeding material and potential cultivars are commonly done in the kinds of yield trials dealt with in the previous section. In contrast, in practical animal breeding there are numerous complications to the estimation of breeding values. Individuals may be related, some individuals but not others may have repeated measurements, and there may be systematic environmental differences between groups of individuals such as management practices (Falconer and Mackay 1996). Not surprisingly, from the 1950s onwards, animal breeders and geneticists developed linear models for the prediction of animal breeding values that took into account all of these complications. The outcome was the General Mixed Model which embraces the vast majority of estimation problems encountered in Quantitative Genetics, including the estimation of breeding values of potential parents (Lynch and Walsh 1998). Today, readers will commonly find genetic models in the scientific literature expressed and analyzed in this form, including ones of relevance to plant breeding. It is therefore worth briefly examining the structure of the General Mixed Model. It is mixed because it includes both

random and fixed effects. Random effects are where treatments (e.g. genotypes from a cross) are a random sample from a large population about which we want to make inferences, whereas with fixed effects we are only interested in the actual treatments used (e.g. no fungicide compared with both the full and half the full recommended application). The general mixed model can be written as follows (Lynch and Walsh 1998).

$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}$  ( $\mathbf{b}$  is sometimes used for the estimate of  $\boldsymbol{\beta}$ , and replaces  $\boldsymbol{\beta}$  in the model)

$\mathbf{y}$  is a column vector ( $n \times 1$ ) containing the phenotypic values of a trait (e.g. yield) measured on  $n$  individuals (or plots in yield trial).

The linear model comprises the following:

$\mathbf{X}$  is  $n \times p$  design matrix with elements 1 (effect contributes) or 0 (effect does not contribute)

$\boldsymbol{\beta}$  is  $p \times 1$  column vector of fixed effects; e.g. trial mean followed by experimental treatments

$\mathbf{Z}$  is  $n \times q$  incidence matrix with elements 1 (effect contributes) or 0 (effect does not)

$\mathbf{u}$  is  $q \times 1$  column vector of random effects; e.g. genetic effects, breeding values

$\mathbf{e}$  is  $n \times 1$  column vector of residual deviations independent of random genetic effects

A special case of the model is the general linear model,  $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$ . When  $\mathbf{X}$  is a design matrix of indicator variables (1 or 0), we have ANOVA (analysis of variance) models. Later in this chapter we shall encounter multiple regression models where  $\mathbf{X}$  contains predictor variables which can take on a continuous range of values. Estimates of  $\boldsymbol{\beta}$  are obtained by the method of least squares in which the residual sum of squares is minimized. The method of ordinary least squares assumes that the residual errors are homoscedastic ( $\sigma^2(e_i) = \sigma_e^2$  for all  $i$ ) and uncorrelated ( $\sigma(e_i, e_j) = 0$  for  $i \neq j$ ), whereas in generalized (weighted) least squares the residual errors can be heteroscedastic (some inherently more variable than others) and/or correlated. If the residuals follow a multivariate normal distribution, the ordinary least squares estimate is also the maximum-likelihood estimate.

The means and variances of the component vectors of the mixed model are as follows.

The means (expected values) of  $\mathbf{u}$  and  $\mathbf{e}$  are  $\mathbf{0}$  by definition.

The mean (expected value) of  $\mathbf{y}$  is therefore  $\mathbf{X}\boldsymbol{\beta}$ .

If  $\mathbf{G}$  is the  $(q \times q)$  covariance matrix of  $\mathbf{u}$  and

$\mathbf{R}$  is the  $(n \times n)$  covariance matrix of  $\mathbf{e}$ , and if  $\mathbf{u}$  and  $\mathbf{e}$  are uncorrelated, then

$\mathbf{V}$ , the  $(n \times n)$  covariance matrix of  $\mathbf{y}$ , is  $\mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R}$  (fixed effects do not contribute). ( $\mathbf{Z}'$  is the transpose of  $\mathbf{Z}$ )

In summary, the random variable  $\mathbf{y}$  has mean  $\mathbf{X}\boldsymbol{\beta}$  and variance  $\mathbf{V}$ .

The mixed-model analysis involves prediction of  $\mathbf{u}$  (random effects) and estimation of  $\boldsymbol{\beta}$  (fixed effects),  $\mathbf{G}$  and  $\mathbf{R}$  from  $\mathbf{y}$ ,  $\mathbf{X}$  and  $\mathbf{Z}$ .

Fixed effects are estimated as Best Linear Unbiased Estimators (BLUEs) and random effects are predicted by Best Linear Unbiased Predictors (BLUPs). They are best because they minimize the sampling variance, linear because they are linear functions of  $\mathbf{y}$  (the observed phenotypes) and unbiased because their expected (mean) values are  $\boldsymbol{\beta}$  and  $\mathbf{u}$ , respectively.

$$\text{BLUE}(\boldsymbol{\beta}) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \quad (\mathbf{V}^{-1} \text{ is the inverse of } \mathbf{V})$$

$$\text{BLUP}(\mathbf{u}) = \mathbf{GZ}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

$\mathbf{G}$  and  $\mathbf{R}$  are generally assumed to be functions of a few unknown variance components and need to be estimated prior to the BLUP analysis by ANOVA or by Restricted Maximum Likelihood (REML). BLUP estimates remain unbiased when estimates of genetic variances are used in place of actual values (as is usual), but are no longer guaranteed to be the best of all unbiased linear estimators. Unlike ANOVA (unbiased) estimation of variance components, maximum likelihood estimation can cope with unbalanced experimental designs. In practice, restricted maximum likelihood (REML) is used to eliminate bias in the estimates of variance components that arises when fixed effects are subject to error. REML estimators maximize only the portion of the likelihood that does not depend on the fixed effects. The estimation is done in an iterative fashion from BLUP estimates of random effects assuming that the appropriate variance components are known. Details can be found in the book by Lynch and Walsh (1998). BLUP is primarily used for the identification of individuals with maximum genetic merit in selection programmes and for monitoring the response to selection. A simple example is given in Box 6.2. As the design is balanced, the components of variance can be derived from an ANOVA. The total sum of squares (80) can be partitioned into treatments (16), genotypes (54) and residual (10) with 1, 7 and 7 degrees of freedom so that the mean squares are 16/1, 54/7 and 10/7. The estimated components of variance are therefore 10/7 for the residual variance and 22/7 (54/7–10/7)/2 for the genotypic variance. The variances actually chosen to generate the data were 1 and 3, respectively. On trying to solve the equation for  $\text{BLUE}(\boldsymbol{\beta}) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$ , one finds a matrix without an inverse (the sum of the second and third rows equals the first row):

$$(\mu) \quad (2.0 \quad 1.0 \quad 1.0)^{-1} \quad (160.0)$$

$$(t_1) = 54/56 (1.0 \quad 3.2 \quad -2.2) \quad 7/54 (123.2)$$

$$(t_2) \quad (1.0 \quad -2.2 \quad 3.2) \quad (36.8)$$

One general solution to the problem is to use the method of generalized inverses to obtain estimable functions and their estimators (Searle 1966). The more common solution is to introduce so-called “convenient restraints” or “obvious restrictions”

such as  $t_1 + t_2 = 0$  ( $\mathbf{X}^*, \beta^*$ ), in other words, the effects sum to zero. There was one degree of freedom for treatments and hence one parameter ( $t$ ) is adequate. The solution to the equations is:

$$\begin{aligned}\mu &= 54/(16 \times 7) \times 7/54 \times (160) = 10 \\ t &= 54/(16 \times 7) \times 7/54 \times (10/54) \times 16 \times 54/10 = 1\end{aligned}$$

Hence  $\mu = 10$  and  $t = 1$ , and these were the values used to generate the data.

We can now estimate  $\mathbf{y} - \mathbf{X}^*\text{BLUE}(\beta^*)$  and hence

BLUP( $\mathbf{u}$ ) as $\mathbf{GZ}'\mathbf{V}^{-1}[\mathbf{y} - \mathbf{X}^*\text{BLUE}(\beta^*)] =$			
(g <sub>1</sub> )	(1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0)	(3)	(3.0)
(g <sub>2</sub> )	(0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0)	(3)	(0.5)
(g <sub>3</sub> )	(0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0)	(0)	(0.5)
(g <sub>4</sub> )	(0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0)	(1)	= 44/54 (2.0)
(g <sub>5</sub> )	(0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0)	(0)	(0.0)
(g <sub>6</sub> )	(0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0)	(1)	(-1.5)
(g <sub>7</sub> )	(0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0)	(1)	(-1.5)
(g <sub>8</sub> )	(0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1)	(3)	(-3.0)
		(0)	
		(0)	
		(-2)	
		(-1)	
		(0)	
		(-3)	
		(-2)	
		(-4)	

The 'g' values used to generate the data were 3, 1, 1, 1, -1, -1, -1 and -3. The estimated values differ from these because, by chance, a genotype could be assessed in two plots with above average fertility or two plots with below average fertility. It is because of this uncertainty that the BLUPs are the mean of the two plots (once fixed effects including the trial mean have been removed) multiplied by the heritability of the trait which is  $[2\sigma_g^2/(2\sigma_g^2 + \sigma_e^2)]$ , where  $\sigma_g^2$  is 22/7 and  $\sigma_e^2$  10/7.

In practice the computation of  $\mathbf{V}^{-1}$  is difficult for very large  $\mathbf{V}$  so  $\beta$  and  $\mathbf{u}$  are jointly obtained from mixed model equations (Henderson 1950):

$$\begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{pmatrix} (\beta) = \begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{pmatrix}$$

**Box 6.2: Example of Use of General Mixed Model ( $y = X\beta + Zu + e$ )**

A simple field experiment is done to assess the yields ( $y$  t/ha) of eight genotypes (a random sample of available genotypes) in plots either treated with a standard fungicide application (treatment 1) or left untreated (treatment 2). Treatments ( $t$ ) and genotypes ( $g$ ) are assigned at random to the 16 plots available. Differences in fertility between the plots contribute to the yield differences ( $e$ ). It is assumed that there is no genotype  $\times$  treatment interaction for the purpose of this simple example. The experiment and the results are summarized in the table below. The trial mean ( $\mu$ ) is 10 t/ha. There is no correlation between the effects of genotypes and  $e$ .

Plot/randomization		Treatment (t)	Genotype (g)	Yield = y
1	8	1	1	14
2	9	2	1	12
3	15	1	2	11
4	10	2	2	10
5	13	1	3	11
6	1	2	3	10
7	2	1	4	12
8	6	2	4	12
9	4	1	5	11
10	7	2	5	9
11	12	1	6	9
12	11	2	6	8
13	3	1	7	11
14	16	2	7	6
15	5	1	8	9
16	14	2	8	5

y	X	$\beta$	$X^*$	$\beta^*$	Z	u	e	$X^*\text{BLUE}(\beta^*)$
14	1 1 0	$\mu$	1 1	$\mu$	1 0 0 0 0 0 0 0	$g_1$	$e_1$	11
12	1 0 1	$t_1$	1 -1	$t$	1 0 0 0 0 0 0 0	$g_2$	$e_2$	9
11	1 1 0	$t_2$	1 1		0 1 0 0 0 0 0 0	$g_3$	$e_3$	11
10	1 0 1		1 -1		0 1 0 0 0 0 0 0	$g_4$	$e_4$	9
11	1 1 0		1 1		0 0 1 0 0 0 0 0	$g_5$	$e_5$	11
10	1 0 1		1 -1		0 0 1 0 0 0 0 0	$g_6$	$e_6$	9
12	1 1 0		1 1		0 0 0 1 0 0 0 0	$g_7$	$e_7$	11
12	1 0 1		1 -1		0 0 0 1 0 0 0 0	$g_8$	$e_8$	9
11	1 1 0		1 1		0 0 0 0 1 0 0 0		$e_9$	11
9	1 0 1		1 -1		0 0 0 0 1 0 0 0		$e_{10}$	9
9	1 1 0		1 1		0 0 0 0 0 1 0 0		$e_{11}$	11
8	1 0 1		1 -1		0 0 0 0 0 1 0 0		$e_{12}$	9

(continued)

**Box 6.2** (continued)

y	X	$\beta$	$X^*$	$\beta^*$	Z	u	e	$X^*\text{BLUE}(\beta^*)$
11	1 1 0		1 1		0 0 0 0 0 0 1 0		$e_{13}$	11
6	1 0 1		1 -1		0 0 0 0 0 0 1 0		$e_{14}$	9
9	1 1 0		1 1		0 0 0 0 0 0 0 1		$e_{15}$	11
5	1 0 1		1 -1		0 0 0 0 0 0 0 1		$e_{16}$	9

## Plot Interference

One potential complication in yield trials is plot interference (Kempton 1997). This occurs when plot yield is affected not only by the genotype grown in the plot, but also by the particular genotypes in neighbouring plots. Interference is most likely to occur when the plots are small and not bordered, as in the early generations of a breeding programme. Interference error can lead to a systematic bias in genotype effects that persists across trials and is not reduced by randomization and replication. Interference sometimes affects only the scale of responses, but can affect the ranking of genotypes. Interference arises from competition for resources, such as light, nutrients or moisture. Thus interference between genotypes may be associated with shading due to height differences in crops such as cereals (Jensen and Federer 1964), cassava (Kawano et al. 1978), field beans (Kempton and Lockwood 1984), kale (Bradshaw 1986) and winter oilseed rape (Talbot et al. 1995), where yields of taller genotypes are enhanced and shorter ones reduced. In contrast, in root and tuber crops such as sugar beet (Kempton 1982), swede (Bradshaw 1989) and potato (Connolly et al. 1993; Bradshaw 1994a), an important factor is root and tuber yields of neighbouring plots. Interference may also be caused by the differential spread of plant pests or pathogens between plots. Thus in a disease screening trial, a highly susceptible genotype may act as a secondary source of infection for other genotypes in the trial, so that the effectiveness of more resistant genotypes is underestimated.

An example of interference is provided in Fig. 6.4 and Table 6.3 from Bradshaw (1994a). Five cultivars of potato (Cara, Estima, Pentland Dell, Pentland Javelin and Torridon) were assessed for 2 years (1992 and 1993) in all possible pairings (a competition diallel) in three consecutive drills (rows), such that one member of the pair (the cultivar) was the centre-drill, and the other member (the neighbour) was in the two adjacent drills. Analyses of variance were done on data from the centre-drills with cultivars and neighbours as fixed effects and years and replicates as random effects. For ware yield (45–85 mm) inter-plant competition was present and the rankings of cultivars in pure-stands were different from those under competition. As there was year  $\times$  cultivar and year  $\times$  neighbour interactions, but no cultivar  $\times$  neighbour interactions, the results for the 2 years are shown separately and include the two-sided neighbour effect. It can be seen that Cara and Torridon

						Neighbour
	CCC	CDC	CEC	CJC	CTC	C
	DCD	DDD	DED	DJD	DTD	D
	ECE	EDE	EEE	EJE	ETE	E
	JCJ	JDJ	JEJ	JJJ	JTJ	J
	TCT	TDT	TET	TJT	TTT	T
Cultivar	C	D	E	J	T	Mean

**Fig. 6.4** Example of a competition diallel with three-row plots in potato (Bradshaw 1994a): cultivars are Cara (C), Pentland Dell (D), Estima (E), Pentland Javelin (J) and Torridon (T). The 25 combinations of cultivar and neighbour had different randomizations in each of four replicates in each of 2 years

**Table 6.3** Potato cultivar fresh-weight ware yields (kg/plot) under competition (comp) and in pure-stands (pure), together with cultivar means (direct effects) and estimated two-sided neighbour effects (neighbour); direct + neighbour = predicted pure

	1992				1993			
	Comp	Pure	Direct	Neighbour	Comp	Pure	Direct	Neighbour
Cara	9.91	9.67	9.86	-0.51	14.36	12.21	13.93	-1.86
Dell	9.49	9.53	9.49	0.04	10.38	12.44	10.79	0.79
Estima	9.94	11.23	10.20	0.44	11.72	14.06	12.19	1.11
Javelin	7.84	8.35	7.94	0.67	9.82	11.22	10.10	1.05
Torridon	10.41	9.77	10.28	-0.60	11.86	11.36	11.76	-1.08
SED	0.54	1.08	0.48	0.48	0.54	1.08	0.48	0.48

are particularly strong competitors, reducing the yields of neighbouring cultivars. Their own yields were higher in competition than in pure stands.

In a competition diallel the yield of the centre plot (drill) of genotype A with genotype B as neighbour is modelled as:

$$\begin{aligned} \text{Yield A with B} &= \text{replicate mean} + \text{direct effect A} + \text{interference effect B} \\ &\quad + \text{interaction A with B} + \text{plot error}. \end{aligned}$$

Thus  $g$  in our model for a randomized complete block trial now becomes:

$$g = v_r + 2a_s + 2i_{rs}$$

where  $v_r$  is the direct effect of cultivar  $r$ ,  $a_s$  the effect of cultivar  $s$  as a one-sided neighbour, and  $i_{rs}$  the non-additive interaction of cultivar  $r$  and neighbour  $s$ . Hence pure-stand yields are  $v_r + 2a_r$ .

There are various ways of dealing with interference (Kempton 1997). In the advanced generations of a breeding programme, when the number of potential cultivars has been reduced and more planting material is available, multiple-row plots can be used and the outer rows discarded. There may also be scope to group similar genotypes (e.g. tall cereals or early maturing potatoes) and thus avoid

interference, provided prior knowledge is available about the trait values for the genotypes. Finally, it may be possible to adjust for interference using an appropriate plant covariate (e.g. plant height or plot yield) derived from neighbouring plots. Such interference models can also be used to test for the presence and size of interference effects and hence alert the breeder to potential problems. However, it is desirable to use multiple-row plots for the final stages of yield assessment of new cultivars because modelling and grouping may not be completely effective. In the potato diallel, for example, individual neighbour effects, but not covariate adjustment, markedly improved the correlations between observed and predicted pure-stand yields. Although competitive ability in potatoes is associated with ‘tuber’ yield, other factors must also be involved; but determining individual neighbour effects is not a practical option when assessing a large number of potential cultivars.

## High-Throughput Phenotyping

Advances in high-throughput genotyping mean that our ability to dissect the genetics of quantitative traits such as yield and stress tolerance is now limited by constraints in field phenotyping capability. Hence there is much interest in developing and using high-throughput phenotyping. Fully automated facilities in greenhouses and growth chambers have been developed using robotics, precise environmental control and remote sensing techniques to assess plant growth and performance, such as the Australian Plant Phenomics Facility (<http://www.plantphenomics.org.au/>). However, results from such facilities do not always translate into improved crop performance under field conditions, so there is also a need to develop high-throughput phenotyping for use in the field. Progress to date has been reviewed by Araus and Cairns (2014). They point out that the most successful traits for evaluation, integrate in time (throughout the crop growth cycle) and space (at the canopy level) the performance of the crop in terms of capturing resources (e.g. radiation, water, and nutrients) and using them efficiently. The different methodological approaches used to evaluate these traits in the field fall into three categories: (i) remote sensing and imaging proximal to the crop, (ii) laboratory analyses of samples and (iii) near-infrared reflectance spectroscopy (NIRS) analysis in the harvestable part of the crop. Besides the choice of the most appropriate traits, it is also crucial to determine the key time(s) for their evaluation. The remote sensing and imaging can be classified as spectral, thermal or digital (camera) and requires empirical calibration against the traits of interest. Simultaneous measurements of all plots within a trial can only be done from the air. Advances in aeronautics and sensors now enable high-quality images to be obtained from unmanned airplanes. An example of a laboratory analysis is carbon isotope discrimination ( $\Delta^{13}\text{C}$ ) in dry matter in relation to drought tolerance. In this context, it should be pointed out that root phenotyping under field conditions remains difficult, and despite all of the limitations, root architecture is still commonly evaluated in clear plastic pots in glasshouse screens. NIRS is now regularly deployed in harvesting machinery to analyze grain characteristics. Although the precision of indirect estimations may be

lower than those of direct analysis, the rapid, low-cost and non-destructive nature of NIRS may justify its use, particularly as a first screen in the early generations of a breeding programme when thousands of genotypes need to be evaluated. Environmental characterization is essential to facilitate data interpretation; for example, patterns of water availability when doing drought phenotyping. Further developments are required in data management and bioinformatics to sensibly evaluate the volumes of phenotypic data which can be collected. High-throughput phenotyping is likely to be a feature of increasing importance in plant breeding.

## Quantitative Trait Loci (QTLs)

So far we have assumed that quantitative traits are controlled by many underlying genes whose individual effects are small and cumulative, but which nevertheless display all of the properties of genes known from Mendelian genetics. But how many genes are actually segregating in any cross, where are they located on the chromosomes, and what are their functions? None of the purely biometrical methods of analysis can give us reliable estimates of gene number and location. However, the discovery of (usually) abundant naturally occurring, single locus polymorphisms at the DNA level, from the late 1970s, opened up the possibility of locating Quantitative Trait Loci through their linkage to these molecular markers. Ever more sophisticated methods of analysis, along with appropriate computer software, have been developed since the 1970s (Liu 1998). An introduction to the methods can be found in *The Genetical Analysis of Quantitative Traits* by Kearsey and Pooni (1996) and a review of relevance to Plant Breeding in *Molecular Plant Breeding* by Xu (2010). Methods are available for analyzing the different types of segregating populations generated by breeders and geneticists, but a simple backcross population is sufficient for my purpose in this chapter.

## Backcross Population: Graphical Genotyping

Let us assume that we have a chromosome 50 cM long, with six evenly spaced molecular markers (*A* to *F*) which have been mapped, and that we are able to make the following backcross:

$$\begin{array}{ccccccccc}
 \text{A} & - & \text{B} & - & \text{Q} & - & \text{C} & - & \text{D} & - & \text{E} & - & \text{F} & \times & \text{abcdef} \\
 & & a & - & b & - & q & - & c & - & d & - & e & - & f & & \text{abcdef}
 \end{array}$$

0    10    15    20    30    40    50 cM

Let us also assume that the yield of the  $\text{F}_1$  ( $\text{ABCDEF}/\text{abcdef}$ ) is 10 t/ha and that of the inbred parent ( $\text{abcdef}/\text{abcdef}$ ) is 6 t/ha, and that this yield difference is due to a single QTL ( $Q/q$ ) located halfway between *B* and *C* at 15 cM.

Furthermore, let us assume that the offspring result from one crossover occurring at random along the pair of ‘female’ chromosomes, so that it is just as likely to occur between *A* and *B* as between *B* and *C*, etcetera. The result will be two recombinant chromosomes and two non-recombinant ones (no crossing-over). Hence 25 % of offspring will be *ABCDEF/abcdef*, 25 % *abcdef/abcdef* and 50 % will have a recombinant maternal chromosome along with a paternally derived *abcdef* chromosome. The observed maternally derived chromosomes, their frequency (%) and yields (t/ha) of the offspring will be as follows:

Chromosome	Frequency	Yield	Chromosome	Frequency	Yield
<b>Non-recombinants</b>					
<i>abqcd</i>	25	6	<i>ABQCDEF</i>	25	10
<b>Recombinants</b>					
<i>Abqcd</i>	5	6	<i>aBQCDEF</i>	5	10
<i>ABq/Qcdef</i>	5	8	<i>abQ/qCDEF</i>	5	8
<i>ABQCdef</i>	5	10	<i>abqcDEF</i>	5	6
<i>ABQCDef</i>	5	10	<i>abqcdEF</i>	5	6
<i>ABQCDEf</i>	5	10	<i>abqcdeF</i>	5	6

*ABq/Qcdef* is the average of *ABqcd* (frequency 2.5 %, yield 6) and *ABQCdef* (frequency 2.5 %, yield 10) and *abQ/qCDEF* is the average of *abQCDEF* (frequency 2.5 %, yield 10) and *abqcDEF* (frequency 2.5 %, yield 6). From this combined graphical genotyping and phenotyping one would correctly conclude that there was a QTL between markers *B* and *C* and that the difference between *Qq* and *qq* was 4 t/ha. To reach this correct conclusion it is important that all of the offspring are correctly genotyped and that the phenotyping is as accurate as possible. With environmental variation and genetic segregation from other chromosome pairs, the phenotypes of the relatively few critical recombinants could vary quite a bit by chance. Hence adequate replication and sample size are important. This simple example allows us to examine the theory behind two of the standard types of QTL analysis.

## Backcross Population: QTL Analysis

If the mean yields of all offspring belonging to the two marker classes at each locus are determined, one obtains the following result:

	<i>Aa</i>	<i>Bb</i>	<i>Cc</i>	<i>Dd</i>	<i>Ee</i>	<i>Ff</i>
	9.4	9.8	9.8	9.4	9.0	8.6
	<i>aa</i>	<i>bb</i>	<i>cc</i>	<i>dd</i>	<i>ee</i>	<i>ff</i>
	6.6	6.2	6.2	6.6	7.0	7.4
Difference	2.8	3.6	3.6	2.8	2.0	1.2

One can then treat the marker genotypes at each locus as classification variables for a t-test, or an analysis of variance, or coded as dummy variables for regression analysis. Whilst it might look as though there is a different QTL associated with each locus, the simplest explanation is a single QTL between *B* and *C*. However, if we had only one marker on this chromosome, *B* for example, we wouldn't know if the QTL was at *B* with effect 3.6 ( $Qq - qq = 3.6$ ), or some distance from *B* with larger effect because:  $Bb - bb = (Qq - qq)(1 - 2r)$  where  $r$  is the recombination frequency between loci *B* and *Q*. In our simple example  $r = 0.05$ , so  $3.6 = (Qq - qq)(0.9)$  and  $Qq - qq = 4.0$ .

However, with a number of marker differences (six in our example), we can regress the marker differences onto  $(1 - 2r)$  and estimate  $(Qq - qq)$  as the slope of the regression line for different possible positions of the QTL along the chromosome. The residual sums of squares for the regressions will be at a minimum (zero in our artificial example) at the true position of the QTL (15 cM in our example), and the slope for this regression will correctly estimate  $(Qq - qq)$  as 4.0. This solution to the problem is known as marker regression analysis.

Kearsey and Pooni (1996) explain how additive ( $a_Q$ ) and dominance ( $d_Q$ ) effects at a QTL locus can be estimated from an  $F_2$  population in which the QTL (*Q*) is linked to a marker (*M*) with recombination frequency  $r$ . If the marker segregates *MM*:*Mm*:*mm* in a 1:2:1 ratio and the means of these marker classes are *MM*, *Mm* and *mm*, then:

$$\frac{1}{2}(MM - mm) = (1 - 2r)a_Q \text{ and } Mm - \frac{1}{2}(MM + mm) = (1 - 2r)^2 d_Q,$$

where  $a_Q = \frac{1}{2}(QQ - qq)$  and  $d_Q = Qq - \frac{1}{2}(QQ + qq)$

Kearsey and Pooni (1996) advocate locating the QTL by regression on to the additive difference between the means and then estimating the dominance effect at this position. In contrast, with interval mapping, multiple regression can be used to estimate  $a_Q$  and  $d_Q$  jointly.

## ***Interval Mapping***

A second method for locating the QTL and estimating its effect is ‘interval mapping’ which uses flanking markers and individual observations rather than marker means. This can be accomplished by maximum likelihood or weighted least squares (regression approach), the two methods giving essentially the same results. Now we start by hypothesizing a QTL between *X* and *Y*, which we will subsequently replace by *A* and *B* etcetera.

$$\begin{array}{c} X - Q - Y \\ x - q - y \end{array}$$

The known recombination frequency between  $X$  and  $Y$  is  $r$  and the recombination frequency between  $X$  and  $Q$  is  $r_x$  and between  $Q$  and  $Y$  is  $r_y$ .

The  $F_1$  will produce six gametes and hence six unique backcross progeny when there is only one crossover between  $A$  and  $F$  (i.e. complete interference so that  $r = r_x + r_y$ ) with the expected frequencies and genetic values ( $Qq = m + \delta$  and  $qq = m$  so that  $Qq - qq = \delta$ ) below:

Genotype	Frequency	Genetic value
$XQY/xqy$	$\frac{1}{2}(1 - r_x - r_y)$	$Qq = m + \delta$
$xqy/xqy$	$\frac{1}{2}(1 - r_x - r_y)$	$qq = m$
$Xqy/xqy$	$\frac{1}{2}r_x$	$qq = m$
$XQy/xqy$	$\frac{1}{2}r_y$	$Qq = m + \delta$
$xQY/xqy$	$\frac{1}{2}r_x$	$Qq = m + \delta$
$xqY/xqy$	$\frac{1}{2}r_y$	$qq = m$

The four marker genotypes are:

Genotype	Frequency	Genetic value
$XY/xy$	$\frac{1}{2}(1 - r)$	$m + \delta$
$xy/xy$	$\frac{1}{2}(1 - r)$	$m + \delta \times 0 = m$
$Xy/xy$	$\frac{1}{2}r$	$m + \delta r_y/r$
$xY/xy$	$\frac{1}{2}r$	$m + \delta r_x/r$

This time the regression is

$$y = \alpha + \beta x$$

where  $y$  is the trait score (10 or 6 t/ha) of the genotype ( $Xy/xy$  can be 10 or 6 depending on whether it is  $XQy/xqy$  or  $Xqy/xqy$  and likewise  $xY/xy$ );  $\alpha = m$ ;  $\beta = \delta$ ; and  $x$  is the coefficient of  $\delta$  ( $1, 0, r_y/r, r_x/r$ ).

Again we test for a QTL at arbitrary positions along the chromosome, using the flanking markers for the position, and locate the QTL at the position where the residual SS is at a minimum. In practice, the model would include a residual for  $j$ th individual,  $\bar{e}_j$ , or epsilon  $j$ . The calculations are done for our simple example in Box 6.3 where the QTL is correctly located midway between markers  $B$  and  $C$  and its size is correctly estimated as 4.00.

### Box 6.3

Interval mapping for six evenly spaced markers  $A$  to  $F$  from 0 to 50 cM along chromosome with QTL at 15 cM and yields of 10 t/ha for  $Qq$  and 6 t/ha for  $qq$  where  $X$  and  $Y$  are the markers flanking  $Q$ : regressions are done for putative QTL at positions 5 (flanking markers  $A$  and  $B$ ), 15 ( $B$  and  $C$ ), 25 ( $C$  and  $D$ ),

(continued)

**Box 6.3** (continued)

35 ( $D$  and  $E$ ) and 45 ( $E$  and  $F$ ) cM along the chromosome so that markers  $X$  and  $Y$ ,  $r$  apart, have QTL midway between them with  $r_x/r = 1/2$  and  $r_y/r = 1/2$

Genotype	Trait score	Coefficient of $\delta$	Frequency (%) of genotypes for flanking markers				
			AB	BC	CD	DE	EF
$XY$	6	1	2.5	0.0	2.5	7.5	12.5
$XY$	10	1	42.5	45.0	42.5	37.5	32.5
$xy$	6	0	42.5	45.0	42.5	37.5	32.5
$xy$	10	0	2.5	0.0	2.5	7.5	12.5
$Xy$	6	$1/2$	5.0	2.5	0.0	0.0	0.0
$Xy$	10	$1/2$	0.0	2.5	5.0	5.0	5.0
$xY$	6	$1/2$	0.0	2.5	5.0	5.0	5.0
$xY$	10	$1/2$	5.0	2.5	0.0	0.0	0.0
$\beta$			3.56	4.00	3.56	2.67	1.78
Residual SS			116	40	116	240	329

The residual SS is at a minimum of 40 with flanking markers  $B$  and  $C$  so the QTL is correctly located equidistant from them, and the size of the QTL effect is correctly estimated as 4.00.

## The Problem of Two Linked QTLs

Let us suppose that there are two QTLs ( $Q_1$  between  $B$  and  $C$  and  $Q_2$  between  $D$  and  $E$ ) of equal effect so that  $Q_1q_2$  and  $q_1Q_2$  both yield 8 t/ha, but that we assume only one QTL. The yields are now those shown below; again  $q/Q$  and  $Q/q$  mean the average of  $Q$  and  $q$ .

Chromosome	Frequency	Yield	Chromosome	Frequency	Yield
Non-recombinants					
$abq_1cdq_2ef$	25	6	$ABQ_1CDQ_2EF$	25	10
Recombinants					
$Abq_1cdq_2ef$	5	6	$aBQ_1CDQ_2EF$	5	10
$ABq_1/Q_1cdq_2ef$	5	7	$abQ_1/q_1CDQ_2EF$	5	9
$ABQ_1Cdq_2ef$	5	8	$abq_1cDQ_2EF$	5	8
$ABQ_1CDq_2/Q_2ef$	5	9	$abq_1cdQ_2/q_2EF$	5	7
$ABQ_1CDQ_2Ef$	5	10	$abq_1cdq_2eF$	5	6

From this combined graphical genotyping and phenotyping one would not necessarily correctly conclude that there were two QTLs contributing to the difference of 4 t/ha between chromosomes because the phenotypic classes might not be clearly distinct. However, on selfing  $ABQ_1Cdq_2ef$ , one could select and test

$ABQ_1Cdq_2ef/ABQ_1Cdq_2ef$  (yield 8 t/ha) and likewise  $abq_1cDQ_2EF$ , which would suggest that each half of the chromosome was contributing 2 t/ha to the yield difference. Further backcrossing followed by selfing would allow finer dissection of the chromosome, with the sections identified by the molecular markers.

If, however, we were to simply rely on marker regression we would conclude that there was a single QTL at 25 cM (residual SS at a minimum), with a difference of 3.0 between  $Qq$  and  $qq$ . However, the fact that the estimated difference is less than the marker differences at 20 and 30 cM, combined with a non-zero residual SS, would suggest a problem with our model. Likewise, if we were to simply rely on the use of flanking markers for ‘interval mapping’ (Box 6.4), we would come to the same conclusion about a single QTL at 25 cM (residual SS at a minimum), but estimate the difference between  $Qq$  and  $qq$  as 3.56.

Fitting the correct two locus model would of course identify two QTL of equal effect, but it would be tempting to incorrectly conclude that there was just one QTL of relatively large effect, particularly with few or poorly situated markers and small population sizes. If on the other hand there were two QTLs of equal effect, but in repulsion rather than in coupling, their effects could cancel out and one would fail to detect a QTL.

**Box 6.4: Interval Mapping when Two QTLs Present, but One QTL Assumed in Analysis**

Genotype	Trait score	Coefficient of $\delta$	Frequency (%) of genotypes for flanking markers				
			AB	BC	CD	DE	EF
$XY$	6	1	2.5	0.0	0.0	0.0	2.5
$XY$	8	1	10.0	7.5	5.0	7.5	10.0
$XY$	10	1	32.5	37.5	40.0	37.5	32.5
$xy$	6	0	32.5	37.5	40.0	37.5	32.5
$xy$	8	0	10.0	7.5	5.0	7.5	10.0
$xy$	10	0	2.5	0.0	0.0	0.0	2.5
$Xy$	6	$\frac{1}{2}$	5.0	2.5	0.0	0.0	0.0
$Xy$	8	$\frac{1}{2}$	0.0	2.5	5.0	2.5	0.0
$Xy$	10	$\frac{1}{2}$	0.0	0.0	0.0	2.5	5.0
$xY$	6	$\frac{1}{2}$	0.0	0.0	0.0	2.5	5.0
$xY$	8	$\frac{1}{2}$	0.0	2.5	5.0	2.5	0.0
$xY$	10	$\frac{1}{2}$	5.0	2.5	0.0	0.0	0.0
$\beta$			2.67	3.33	3.56	3.33	2.67
Residual SS			160	70	36	70	160

When two QTLs (linked or unlinked) have been detected using QTL mapping, a two-QTL model can be constructed and tested using the contrasts on the QTL genotype means to detect and estimate the effects of non allelic interactions as well

as those at individual loci. Quantitative resistance to *Globodera pallida*, the white potato cyst nematode, provides three examples of two QTLs acting additively to give a high level of resistance to this major worldwide pest of potato. QTLs have been found on chromosomes 5 and 9 for *S. vernei*-derived resistance (Rouppé van der Voort et al. 2000; Bryan et al. 2002), on chromosomes 4 and 11 for *S. tuberosum* subsp. *andigena*-derived resistance (Bryan et al. 2004), and on chromosomes 5 and 11 for *S. sparsipilum*-derived resistance (Caromel et al. 2005). Hence there would appear to be very good prospects of combining at least three QTLs on different chromosomes to give a very high level of resistance which may prove durable. Molecular markers would be very valuable to ensure that three different QTLs are present in the same potato genotype (potential cultivar).

## Marker-Assisted Selection Versus Phenotypic Selection in Winter Wheat

This would seem to be a good point to pause and look at a comparison done by Miedaner et al. (2009) of phenotypic and marker-assisted selection for *Fusarium* head blight resistance in winter wheat, based on three quantitative trait loci (QTL) from two European sources. Each QTL allele explained about 20 % of the variation in the original mapping populations. *Fusarium* head blight is a major disease of wheat, causing yield losses and mycotoxin contamination. Resistant cultivars are highly desirable and Miedaner et al. (2009) concluded that resistance QTLs already present in agronomically improved winter wheat germplasm were an attractive proposition for the high-yielding areas of NW Europe. They decided to use two QTLs from 'Dream' and one from 'G16-92' in a double-cross with two high-yielding, susceptible winter wheat's, 'Brando' and 'LP235.1', in such a way as to combine the three QTLs which were on different chromosomes [(Dream × Brando) × (G16-92 × LP235.1)]. The double-cross population comprised 600 individuals. Sixty plants possessing SSR markers closely linked to the QTL alleles were selfed, harvested separately, and 60 offspring of each (3600 in total) analysed for the presence of the markers. Thirty five offspring possessing either all three (17) or two out of three ( $9 + 9 = 18$ ) markers in the homozygous state were selected and selfed again. The resulting 35 lines formed the marker selected population (CM).

All 600 double-cross plants were in fact selfed so that 600 lines could be assessed for resistance at four locations with two replicates after inoculation with *F. culmorum*. Twenty lines with the best resistance ratings were selected and intercrossed in a greenhouse: each line was crossed with three other lines, using four plants of each, making 240 ( $20 \times 3 \times 4$ ) crosses in total. Five plants from each cross were selfed to produce 1200 lines of which 1000 (phenotypically selected population CP) were tested in the field at four locations with two replicates. Also included in these field assessments were a random sample of 50 out of the 600 lines (unselected control population CO) and the 35 marker selected lines. The population means for CO, CP and CM were 23.6 %, 17.4 % and 18.6 % infection, respectively.

Heritabilities of line means (across environments and replicates) were high for CP (0.91) and CM (0.94). Realized selection gain was calculated as the difference between the unselected base population (CO) and the selected populations CP and CM. The reductions in disease severity were 6.2 and 5.0, respectively; but 2.1 and 2.5 per year because marker selection saved a year. The marker and phenotypic selection schemes were designed to correspond to practical wheat breeding. Marker-selected lines were on average 8.6 cm taller than phenotypically selected lines. Genotypic variation within marker-classes was large and many lines with high resistance were found in the CM population. Hence after marker selection to combine QTL alleles for resistance, phenotypic selection would seem desirable to achieve further gains in resistance, and possibly some reduction in plant height.

## Permutation Tests

Before continuing with methods of QTL analysis, we need to pause and consider the issue of when to declare an apparent QTL statistically significant. Let us start with a single-marker analysis of a trait and determine the difference in mean scores between the two marker classes, presence and absence of marker, for a very large number of markers. We can determine the significance of each difference by a *t*-test. However, if we noted all differences significant at the 5 % level, then with a large number of markers, 5 % would be significant by chance. A more appropriate threshold value for declaring a significant QTL effect can be found through the empirical method of Churchill and Doerge (1994). The trait scores are shuffled and analysed for QTL effects at all of the markers and the largest *t*-value is stored. The entire procedure is then repeated 1000 times. The stored *t*-values are then ordered and the 95 % quantile taken as the estimated critical value at which the overall type-I error rate for the experiment is 0.05 (5 %) or less. A type-I error is when we accept a chance result as a real effect (type-II error is when we reject a real effect as chance). The method is a general one for determining appropriate significance thresholds (critical values) against which to compare test statistics (usually LOD scores or likelihood ratios) for the purpose of detecting QTL. It can therefore be used with any QTL mapping procedure including simple linear regression, multiple regression and multiple regression with cofactors (see next section). It overcomes the problem of determining (or approximating) the distribution of the test statistic under an appropriate null hypothesis, and also the multiple hypothesis testing that is implicit in the genome searches used for locating QTL (a large number of tests may be carried out, many of which are not independent). As the procedure is empirical, based on the observed marker and trait data, it will automatically reflect the characteristics of the particular experiment to which it is applied. One word of caution is, however, required. As QTL are reported only when the test statistic reaches the predetermined critical value, the estimated effects of detected QTL are actually sampled from a truncated distribution. As a consequence, the estimated number of QTLs of small to intermediate effect (each explaining 0.75–9.5 % of

phenotypic variation) will be biased downwards, but the average effect of the detected ones will be biased upward; a phenomenon known as the Beavis effect (Xu 2003). With QTL of small effect, there is little bias when 1000 progeny are evaluated, slight over estimation with 500, but serious over estimation with only 100. Utz et al. (2000) recommended Cross-Validation for obtaining asymptotically unbiased estimates of the proportion of genotypic variance explained by QTL and consequently a realistic (rather than optimistic) assessment of the prospects of marker assisted selection. They evaluated testcrosses of 344 F<sub>3</sub> maize lines grown in four environments for a number of agronomic traits. They divided their data sets into ones for mapping QTLs and estimating the genetic variance explained and others for independent validation of the latter. The proportion of genotypic variance (and hence QTL effects) explained in the test sets was only 50 % of that found in the estimation sets; in other words, the latter were biased. Fivefold cross-validation is explained in more detail later in this chapter under genomic selection.

## Composite Interval Mapping

Composite interval mapping (CIM) has been developed to increase QTL resolution over that achieved by simple interval mapping, and hence potentially to resolve the number of segregating QTLs in a linkage group (Jansen 1993; Zeng 1993, 1994). A detailed account can be found in the book by Liu (1998). CIM is a combination of simple interval mapping and multiple regression in which the simple model is extended to contain the partial regression coefficients ( $\beta$ 's) of the trait value on markers (e.g. Z/z) other than the flanking markers so that the x's are dummy variables with values 1 and 0 for Zz and zz, respectively, in a backcross. The markers in the CIM model can control the residual genetic background when they are linked to other QTLs, thus leaving environmental variation that can be controlled through the experimental design (e.g. more replication). In practice one needs to identify a small set of markers which are close to possible QTLs. In the barley example given by Liu (1998), five markers were used on the chromosome where an interval with flanking markers was being considered, and one marker on each of four other chromosomes. The result was a higher resolution of the QTL between the flanking markers and the detection of two other QTLs on the same chromosome. The markers for inclusion in the model can be found by stepwise multiple regression analysis; and in effect this is a strategy for detecting multiple QTLs segregating for a trait in a population. The first pass through the data finds the QTL with largest effect which is then included in the model for the second pass to find the next QTL, and so on until no more QTLs are detected. However, this does not mean that all of the QTLs have been found, or that the undetected ones of small effect don't matter to a plant breeder. Searching the whole genome simultaneously for multiple QTLs might seem a better strategy, but the information content of typical experiments is inadequate. Theoretical (power) calculations suggest that the upper limit is 9–12 for the number of QTLs detectable in typical experiments

(Kearsey and Pooni 1996). Common practice has been to use population sizes of around 100, but 200 would be better. Furthermore, the 95 % confidence interval for a QTL (95 % chance of QTL actually being in interval) can be large, as reported by Kearsey and Pooni (1996): computer simulation has shown that with an  $F_2$  population of 300 individuals, five well-spaced markers and a QTL with heritability 20 %, the 95 % confidence interval spans 30 cM, i.e. 30 % of a typical chromosome. In order to reduce this interval significantly, populations of several thousand individuals need to be genotyped and scored. It is important to emphasize that population size and heritability, but not marker density to any appreciable extent, determine the number and size of QTL that can be detected. This is because crossing-over is a relatively rare event and mapping depends on recombination. The difficulty in separating two linked QTL from one is therefore not surprising, and can still remain despite greater resolution of QTLs. Hence the attraction of using markers to construct chromosomes containing short, prescribed regions from different sources in order to locate QTLs and determine their effects. An interesting example is introgression mapping in *Lolium/Festuca* grasses because it can be combined with the transfer of desirable alleles (genes) from one species to another.

## Introgression Mapping in *Lolium/Festuca* System

King et al. (1998) crossed a synthetic autotetraploid *Lolium perenne* (*Lp* variety Meltra,  $2n = 4x = 28$ ), as the female parent, with diploid *Festuca pratensis* (*Fp* Bf 1183,  $2n = 2x = 14$ ) as the pollen parent. A single *LpLpFp* triploid hybrid ( $2n = 3x = 21$ ) from this cross was then used as the pollen parent in crosses to a single diploid *Lp* genotype (variety Liprio). From the resulting backcross ( $BC_1$ ) population of 161 plants, an initial sample of 78 was selected and screened using GISH, as the chromatin of the two species can be distinguished by this technique. The screening revealed that 75 had 14 chromosomes and three had 15. Fifty-eight plants, including the three with 15 chromosomes, carried one or more *Fp* chromosomes or chromosome segments. Three lines carried *Lp/Fp* chromosome substitutions and these were fully fertile near-normal plants, suggesting the substitution of a homoeologous chromosome. Fourteen plants carried a single *Fp* chromosome segment which ranged in size from a small segment to more than a whole chromosome arm. Thirty-one plants had two *Fp* chromosome segments, nine had three and one had six. The remaining 20 plants analysed had no *Fp* chromosome segment detectable by GISH. By comparing the physical size of introgressed *F. pratensis* chromosome segments with the presence or absence of *F. pratensis*-specific molecular markers (AFLPs, RFLPs and RAPDs), it was possible to determine the physical position of genetic markers. Analysis of the *Lp/Fp* recombinant chromosomes revealed that homoeologous recombination in the parental *LpLpFp* triploid plant appeared to occur along the entire length of the *Lp* and *Fp* chromosomes. Nevertheless, further analysis did reveal that preferential pairing occurred between the *Lp* homologous chromosomes in the *LpLpFp* triploid, but that *Fp* chromosomes paired

homoeologously with their *Lp* partners at equal frequency (King et al. 1999). Plants identified as carrying a single *Fp* chromosome using GISH were backcrossed again as male parent to *Lp* (variety Liprio) as female parent to produce a second backcross ( $BC_2$ ) population. This allowed further dissection of *Fp* chromosome 3 into 18 physically demarcated sections (bins). In addition, mapping of 148 back-cross individuals in one  $BC_2$  population with 104 AFLPs generated a marker map of 81 cM, compared with a predicted length of 76 cM from chiasma frequency counts (King et al. 2002). Hence there was a 1:1 correspondence between chiasma frequency and recombination and, in addition, the absence of chromatid interference across the *Festuca* and *Lolium* centromeres. Overall, this research demonstrated that introgression mapping could be used to align physical and genetic maps, determine the genetic control of agronomically important characters, and produce novel germplasm for the development of new cultivars.

Subsequently, King et al. (2007) were able to exploit the published rice genome sequence to compare the synteny between rice chromosome 1 and *Lolium/Festuca* chromosome 3. Without going into details, they were in effect able to use PCR primers to amplify the same 69 DNA (BAC/PAC clones) conserved sequences in rice and *Lolium/Festuca* and assign SNPs derived from rice linkage group 1 to bins on the *Festuca* introgression map. They demonstrated that a substantial component of the coding sequences in monocots is localized proximally in regions of very low and even negligible recombination frequencies. The implications are that during domestication of monocot plant species selection has concentrated on genes located in the terminal regions of chromosomes within areas of high recombination frequency, and that a large proportion of the genetic variation available for selection of superior plant genotypes has not been exploited.

## Chromosome Segment Substitution Lines (CSSLs)

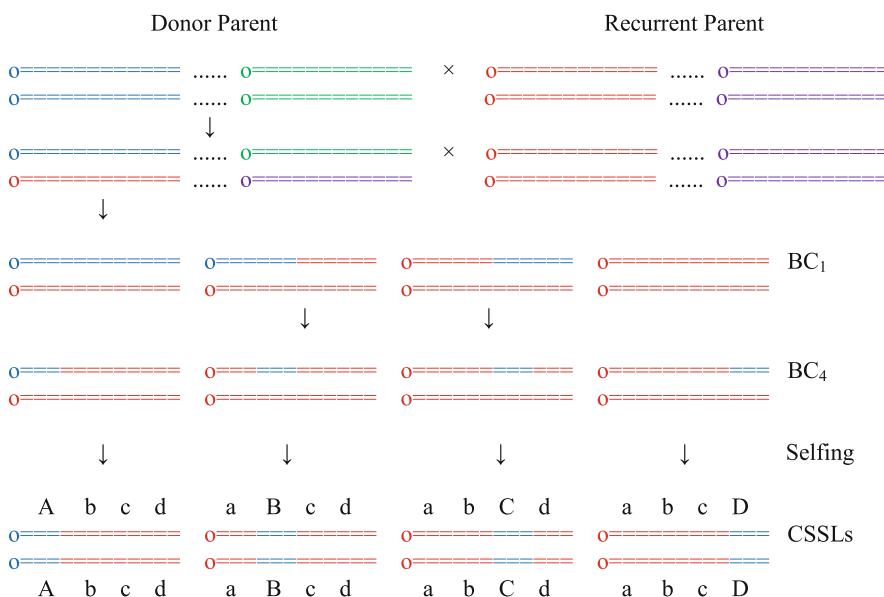
The production of chromosome segment substitution lines by backcrossing, like the *Lolium/Festuca* system, combines QTL mapping with the production of novel germplasm for the development of new cultivars. Three examples will be considered, one from peanut and two from rice. It is of most use with inbreeding species. The theory is shown in Table 6.4 and Fig. 6.5. Let us start with a simple reminder of backcrossing. If the donor parent has allele *A* at a locus and the recurrent parent has allele *a*, after four backcrosses ( $Aa \times aa; \times aa; \times aa; \times aa$ ) there is a 1 in 16 chance that an individual will be *Aa*. If we consider eight unlinked loci, *A* to *H*, the same is true for each locus and we can use the binomial distribution to work out the probabilities of an individual having no donor alleles, one donor allele, etc. (Table 6.4). Looked at another way, if we have a sample of eight individuals, on average 4.8 will have no donor alleles, 2.5 will have one allele, 0.6 will have two and 0.1 will have three. If these heterozygote individuals are selfed for two generations, we will be able to identify eight true breeding lines with these numbers of donor alleles. We are thus led to the idea that it should be possible to produce eight lines

**Table 6.4** Chromosome segment substitution lines

Segments	Probability	Number of lines
0	$Q^8$	0.5967
1	$8PQ^7$	0.3183
2	$28P^2Q^6$	0.0743
3	$56P^3Q^5$	0.0099
4	$70P^4Q^4$	0.0008
5	$56P^5Q^3$	0.0000
6	$28P^6Q^2$	0.0000
7	$8P^7Q$	0.0000
8	$P^8$	0.0000

Probabilities of 0–8 donor segments in a sample of eight lines (from binomial distribution)

After four backcrosses, probability of one segment is  $P = 1/16$ , ( $Q = 1 - P$ )



**Fig. 6.5** Production of chromosome segment substitution lines (blue to green is donor parent and red to purple is recurrent parent in backcrossing): four segments A to D on one chromosome are shown

each with a single unique donor allele. In practice loci can be linked and we are dependent on meioses to produce chromosome segments, as shown in Fig. 6.5. Here the ideal situation of non-overlapping fragments and complete genome coverage is shown. In practice there will be overlap of segments and possibly incomplete genome coverage. Nevertheless, molecular markers can be used to select the desired fragments and reject the undesirable ones, and hence make the construction of CSSLs an efficient process. Each line will therefore consist of one or few

homozygous chromosome segments derived from a donor parent in the genetic background of a recurrent parent. The effect of each chromosome segment on a trait can then be evaluated relative to the recurrent parent. The donor genome is therefore dissected, but without assessing genetic interactions among QTLs. As the lines are genetically fixed, they can be simultaneously and repeatedly assessed in diverse environments for many traits. The size of 'QTL' effect that can be detected will depend on the design of the experiment, as explained at the beginning of this chapter. With adequate replication and control of environmental variation through blocking and spatial analysis, quite small effects can be detected.

## Peanut

Cultivated peanut (*Arachis hypogaea*) is a tropical legume that is native to South America. It is an allotetraploid ( $2n=4x=40$ ; AB genome) with limited genetic diversity. Hence the introgression of desirable genes from its wild progenitors (and other wild relatives) is an important objective in many peanut breeding programmes. Fonceka et al. (2012) used a marker-assisted (115 SSRs, average spacing 12 cM) selection (MAS) backcrossing strategy to produce a population of 122 CSSLs from the cross between the wild synthetic allotetraploid [AiAd: *A. ipaënsis* KG30076 (diploid BB genome)  $\times$  *A. duranensis* V14167 (diploid AA genome)] and cultivar Fleur11 as recurrent parent. Four backcross generations (22/88, 58/192, 80/565 and 100/807 individuals by MAS) were followed by two generations of selfing (140/1180 and 122/1078). Fleur 11 was used as female parent for the first two backcrosses and then as male parent for the next two backcrosses. The aim of the MAS was to identify a minimum subset of individuals providing an optimal coverage of the cultivated peanut genome with chromosome segments from the wild donor. The 122 CSSLs were on average 97.28 % homozygous for Fleur 11 and 2.24 % homozygous for AiAd genes. The mean number of introgressed segments per line was 1.5 and the mean size of a segment was 39.24 cM (range 5.7–115.6 cM). Seventy two lines carried one unique donor segment, 37 lines carried two segments, ten lines carried three segments, one line carried four segments and two lines carried five segments. As a demonstration of the utility of these lines, a subset of 80 were evaluated in a replicated field trial in Bambey (14.42°N and 16.28°W), Senegal for aspects of peanut morphology. The introgressed segments of the 80 CSSLs covered 88.7 % of the genetic map. A total of 28 lines showed significant differences from Fleur11. Among these lines, 15 had one unique wild chromosome segment and 11 had two wild segments. When a given line had more than one wild segment, comparisons were made with lines having overlapping segments and similar phenotypes. As a result, 42 QTLs were detected for all traits: 14 for plant growth habit, 15 for height of the main stem, 12 for plant spread and one for flower colour. Among the 42 QTLs, 37 were assigned to genomic regions and three QTL positions were considered putative. More details can be found in the paper. Further assessment of the lines is required to determine their value for

breeding. More backcrossing could also be done to produce a set of CSSLs with complete genome coverage with one unique wild chromosome segment in each line. This peanut example has provided a clear example of the fact that whilst a breeder cannot currently control what goes on during meiosis, molecular markers can be used to monitor and select the desired products for further breeding and genetic analysis.

## Rice

The wild relatives of rice are useful sources of alleles that have been selected for adaptation to diverse environments around the world. *Oryza rufipogon*, the known progenitor of cultivated rice (*Oryza sativa* L.), contains alleles that were not actively selected during domestication, and hence lost, making it an ideal source of genes that can be utilized to improve existing rice cultivars; for example for biotic and abiotic stress resistance. Furuta et al. (2014) produced 33 chromosome segment substitution lines (CSSLs) of *O. rufipogon* (W0106 from India) in the background of the elite *japonica* cultivar Koshihikari. As in the peanut example, at least four backcrosses (87–172 individuals) with cultivar Koshihikari as the male parent were followed by two generations of selfing, again using marker assisted selection (149 SNPs). Over 90 % of the entire genome was introgressed from the donor parent into the CSSLs. A total of 99 putative QTLs were detected in field assessments in Japan, of which 15 had large effects on the traits examined (days to heading, culm height and panicle and grain characteristics). Three of these QTLs showed agronomically positive effects on the traits, two on grain length, and one on chromosome 10 which increased the number of grains per panicle by 46.9 %. Comparison of the putative QTLs identified in this study and previous ones indicated a wide genetic diversity between *O. rufipogon* accessions. Furuta et al. (2014) also mention CSSLs derived by other researchers from distant relatives of rice, including *O. meridionalis*, *O. glumepatula* and *O. glaberrima* as well as *O. rufipogon*.

My second rice example concerns cooking quality. In the best-quality grain, the kernel elongates but its width changes little during cooking, a trait which has been associated with the Basmati type of rice. More generally, greater cooked rice elongation (CRE) is a characteristic of high-quality cultivars. Yang et al. (2013a) produced and assessed in Nanjing (China), 103 chromosome segment substitution lines (CSSLs) of Nipponbare in the background of the elite cultivar 93-11. Four backcrosses to 93-11 were followed by two generations of selfing, again using marker assisted selection (230 SSRs). Each CSSL contained one substituted segment of Nipponbare in the 93-11 genetic background. Milled rice length (MRL) and cooked rice length (CRL) were assessed in addition to cooked rice elongation (CRE). In total, 12 QTLs for rice elongation traits were detected: two for MRL on chromosome 3, one for MRL on chromosome 8, four for CRL on chromosomes 3, 6, 8, and 9, and five for CRE on chromosomes 4, 6, 9, 10, and 11. The authors

concluded that *qCRE-6* located in the region close to the *Wx* (waxy) gene might be important for improving CRE, and that 93-11 could be a desirable parent for improving the quality of rice by introducing its *Wx* alleles into breeding lines.

## Mapping of QTL by Whole Genome Resequencing of DNA from Two Bulked Populations

An interesting method (QTL-seq) for the rapid mapping of QTL was proposed by Takagi et al. (2013). It involves whole genome resequencing of DNA from two bulked populations. Briefly the procedure is as follows. Let us consider 256 true breeding lines from a cross between a tall (7 ft) and short (1 ft) cultivar which segregates for eight unlinked genes of equal effect so that 37 lines (14 %) have a height of 5.5 ft or more and 37 lines have a height of 2.5 ft or less. A sample of DNA is taken from each of these lines and two bulks are made, one of DNA from the tall lines (tall bulk) and the other of DNA from the short lines (short bulk). Whole genome resequencing is done on each bulk with greater than 6x genome coverage. A large number of sequence-reads each of 75 bp is obtained and aligned with a reference map of one or both parents. The result is random sets of  $n$  alleles from each bulk at frequent intervals (loci) throughout the whole genome. One then looks for SNPs across the whole genome. For example, the low parent may have base ‘C’ and the tall one base ‘G’. If the short bulk is a mixture of ‘Cs and Gs’ in roughly a 1:1 ratio (SNP-index of 0.50), and the tall bulk likewise, then there is no association with height. However, if the short bulk contains all ‘Cs’ (its SNP-index is 0.0) and the tall bulk contains all ‘Gs’ (its SNP-index is 1.0) there is an association with height, and the difference in SNP-index (tall–short) at this location will be 1.0. Should the short and tall bulks contain all ‘Gs’ and ‘Cs’, respectively, the difference would be –1. In practice, 10 consecutive SNPs are used to produce ‘a sliding window’ value ( $m$ ) of this difference which is plotted against position along chromosome. QTLs show up as peaks of value 1.0 (or –1.0) from a baseline of 0.0. In fact values of  $m$  greater than about 0.3 were shown to be significant ( $P < 0.01$ ) by simulation. Takagi et al. (2013) demonstrated that the method worked in rice for partial resistance to the fungal pathogen *Magnaporthe oryzae* (cause of rice blast disease) in a population of 241 F<sub>7</sub> recombinant inbred lines (QTL on chromosome 6), and for seedling vigour in an F<sub>2</sub> population of size 531 individuals (QTLs on chromosomes 1 and 3). They also demonstrated by simulation that high and low bulks each comprising 20–50 individuals and 15 % of the population together with an  $n$  value of 20 ‘alleles’ would be a reasonable choice for an F<sub>2</sub> population. Furthermore the method had sufficient power to detect QTLs accounting for 10 % of the total variation.

## Candidate Genes

A problem with the methods just described is that they may be cross specific. A marker-trait association found in one QTL mapping population may not occur in another population and hence the marker is only of limited use to the breeder. In other words, across populations, the marker and QTL allele are in linkage equilibrium (not associated). However, as an increasing number of economically important plant species have their genomes sequenced, and as understanding increases of the biochemical pathways underlying traits of importance, more and more candidate genes for QTL are likely to be identified. One will then be able to develop primers for PCR amplification of DNA fragments for sequencing part, or all, of the candidate gene in a number of individuals (genotypes) from a wide range of germplasm. Associations can then be sought between variations in the DNA sequences and variations in the phenotypic trait. Desirable QTL alleles are likely to be recognized through their single-nucleotide polymorphism (SNP) haplotypes (specific combination of nucleotides at sites showing polymorphism) which can then be manipulated in a breeding programme as though they were Mendelian genes affecting qualitative traits. The limitation of this approach is that the causative mutation may lie in a gene that is not at present an obvious candidate for the trait. An example of this approach is the paper by Skøt et al. (2007) on the association of candidate genes with flowering time and water-soluble carbohydrate content in *Lolium perenne*. This is an important temperate forage and amenity grass that is an obligate out-breeding species in which short days and low temperatures (vernalization) are required as a primary induction of flowering, followed by longer days and higher temperatures. The candidate genes chosen for the two traits were a cytosolic neutral/alkaline invertase gene (*LpcAI*) for water-soluble carbohydrate and a homologue of the *CONSTANS* gene in *A. thaliana* (*LpHDI*) for flowering time (heading date). In total, 96 genotypes from each of nine European populations of *L. perenne* were used in the research, of which seven were natural or seminatural and two were varieties. Analysis of 506 AFLP markers confirmed that the 864 genotypes were clustered in nine groups, coinciding with the nine populations, and this population structure was taken into account when looking for associations. The most interesting result was a potential candidate SNP (4443), identified in the *LpHDI* locus, which was consistently associated with flowering time (heading date). It has promise as a marker for flowering time in turf-grass breeding where one of the goals is to obtain elite cultivars with an earlier flowering time, since this is likely to enhance seed yield as harvesting conditions will be more favourable earlier in the year.

## Expression Quantitative Trait Loci (eQTLs)

In the last chapter microarrays were introduced as a tool for monitoring the gene expression levels of thousands of genes in parallel; that is the abundance of mRNA transcripts. Microarray technology can also be used to obtain genome-wide expression profiling from the individuals in an immortal mapping population such as an RIL population. In future, next generation sequencing may be used. Levels of gene expression are equated with the steady-state abundance of individual mRNA transcripts that have been determined in a specific sample from the individual at a given point in time. This enables the mapping of QTLs controlling the transcript level for each gene (expression or eQTLs) and hence a comparison of the relationship between genome and transcriptome, as reviewed by Hansen et al. (2008) and Druka et al. (2010). The eQTLs can then be used to identify candidate genes from associations between gene expression polymorphisms and a phenotypic (trait) QTL. Furthermore, eQTL mapping data enables genetic regulatory networks to be modelled to provide a better understanding of the underlying phenotypic variation. Expression QTLs are categorized as *cis* or *trans*, where *cis* eQTLs represent a polymorphism physically located near the gene itself; for example, a promoter polymorphism that gives rise to differential expression of the gene. In contrast, *trans* eQTLs result from polymorphisms at a location in the genome other than the actual physical position of the gene whose transcript level is being measured. This region could, for example, contain a polymorphism in the expression of a transcription factor that modulates the transcript level of the target gene(s). A single gene can have one or multiple eQTLs. Druka et al. (2010) explain how microarray experiments can provide genome-wide genetic markers (transcript-derived markers) as well as the transcript abundance phenotype of every gene that corresponds to a probe or probe-set on the array. They make recommendations on methodology based on their experience with barley.

## Association Genetics and Genomic Selection

When a trait is affected by a large number of genes each of small effect, genome wide association (mapping) studies (GWAS) is an attractive approach to identifying markers very tightly linked to the desirable QTL alleles. This is a population approach in which genetic variation is considered at two or more loci simultaneously. Hence allele frequencies at each locus are insufficient to describe the genetic variation. Multilocus gamete frequencies must be utilized because associations of alleles within gametes may occur. Linkage disequilibrium (LD) is the non-random association of alleles at different loci into gametes. A particular sequence of tightly linked alleles (or nucleotides within an allele) is often referred to as a haplotype. The prerequisite for GWAS is the capacity to identify and score tens-of-thousands of markers at low cost, and this is now becoming possible with

high-density single-nucleotide polymorphism (SNP) markers. In other words, multi-locus linkage disequilibrium is sought between QTL alleles and markers across the whole genome in a set of germplasm referred to as the training population. This germplasm may be a diverse set of genotypes from a collection of interest to the breeder (population mapping) or lines in a breeding programme derived from multiple biparental crosses (family mapping). Family mapping is based on the analysis of multiple segregating families, typically from connected crosses. Results on optimal design for family mapping can be found in the paper by Liu et al. (2013), but need not concern us here.

In theory it should be possible to account for the entire genetic component of phenotypic variation in terms of variation in markers tightly linked to all of the segregating QTLs. In other words, it should be possible to determine the genetic architecture of the trait: the number of genetic loci and the allele distribution at each locus. The breeder will then want to use tightly linked markers to practise selection for desirable QTL alleles in newly generated populations and families that have not been phenotyped, and which are referred to as the selection (or testing/validation) population. The hope is that the desirable QTL alleles and same tight linkages will be present in this newly created germplasm. Where relatively few QTLs are found, markers tightly linked to the desirable QTL alleles will be used for marker assisted selection. However, there is now a lot of interest in genomic selection (GS) where markers are used to predict the genetic values of breeding material; that is, genomic estimated breeding values (GEBVs). It is common for all markers to be fitted simultaneously to avoid biased marker effects and to capture all small effects. Genomic selection (GS) has the potential to improve the efficiency of breeding programmes in two ways. Firstly, to reduce the cycle time between one set of crosses and selecting parents from the progeny for the next set of crosses. Secondly, to reduce the amount of phenotypic evaluation in costly field trials conducted over seasons and locations, aimed at identifying potential cultivars. The germplasm may be a set of clones, inbred lines or hybrids.

We can initially explore ideas further in a simple example without asking how the linkage equilibrium and disequilibrium between markers and QTL alleles arose in the first place. We shall see in Chap. 9 that very tight linkage means an ancient association between QTL and marker so that a large number of recombination events have established equilibrium with other markers.

## *A Simple Example*

Consider the set of inbred lines in Box 6.5. Five tightly linked marker loci *A-E*, each with two alleles, are in linkage equilibrium. There is no association between the two alleles at one locus and the two at any other locus. For example, allele *A* is equally likely to be found with *B* and *b*, *C* and *c* etc. The measured trait, say yield, is affected by two QTL of equal and independent effect so that genotypes  $Q_1Q_1Q_2Q_2$ ,  $Q_1Q_1q_2q_2$ ,  $q_1q_1Q_2Q_2$  and  $q_1q_1q_2q_2$  have yields of 10, 8, 8 and 6 t/ha, respectively.

We now determine the yields ( $Y$ ) of the 32 ( $N$ ) inbred lines and their marker allele composition and then perform the regression of phenotypic values on all available markers, as first proposed by Meuwissen et al. (2001). In other words, we do a standard multiple regression analysis.

### Box 6.5

Association genetics: yields of 32 inbred lines (departures from mean of 8) and marker genotypes at five tightly linked loci  $A$ ,  $B$ ,  $C$ ,  $D$  and  $E$  in linkage equilibrium; there are two QTLs ( $Q_1$  and  $Q_2$ ) of equal effect in linkage disequilibrium with  $B$  and  $D$ , respectively, and linkage equilibrium with  $A$ ,  $C$  and  $E$

Line	Yield mean 8	Locus $A$	Locus $B$ and $Q_1$	Locus $C$	Locus $D$ and $Q_2$	Locus $E$
1	+2	$A +\frac{1}{2}$	$BQ_1 +\frac{1}{2}$	$C +\frac{1}{2}$	$DQ_2 +\frac{1}{2}$	$E +\frac{1}{2}$
2	+2	$A +\frac{1}{2}$	$BQ_1 +\frac{1}{2}$	$C +\frac{1}{2}$	$DQ_2 +\frac{1}{2}$	$e -\frac{1}{2}$
3	0	$A +\frac{1}{2}$	$BQ_1 +\frac{1}{2}$	$C +\frac{1}{2}$	$dq_2 -\frac{1}{2}$	$E +\frac{1}{2}$
4	0	$A +\frac{1}{2}$	$BQ_1 +\frac{1}{2}$	$C +\frac{1}{2}$	$dq_2 -\frac{1}{2}$	$e -\frac{1}{2}$
5	+2	$A +\frac{1}{2}$	$BQ_1 +\frac{1}{2}$	$c -\frac{1}{2}$	$DQ_2 +\frac{1}{2}$	$E +\frac{1}{2}$
6	+2	$A +\frac{1}{2}$	$BQ_1 +\frac{1}{2}$	$c -\frac{1}{2}$	$DQ_2 +\frac{1}{2}$	$e -\frac{1}{2}$
7	0	$A +\frac{1}{2}$	$BQ_1 +\frac{1}{2}$	$c -\frac{1}{2}$	$dq_2 -\frac{1}{2}$	$E +\frac{1}{2}$
8	0	$A +\frac{1}{2}$	$BQ_1 +\frac{1}{2}$	$c -\frac{1}{2}$	$dq_2 -\frac{1}{2}$	$e -\frac{1}{2}$
9	0	$A +\frac{1}{2}$	$bq_1 -\frac{1}{2}$	$C +\frac{1}{2}$	$DQ_2 +\frac{1}{2}$	$E +\frac{1}{2}$
10	0	$A +\frac{1}{2}$	$bq_1 -\frac{1}{2}$	$C +\frac{1}{2}$	$DQ_2 +\frac{1}{2}$	$e -\frac{1}{2}$
11	-2	$A +\frac{1}{2}$	$bq_1 -\frac{1}{2}$	$C +\frac{1}{2}$	$dq_2 -\frac{1}{2}$	$E +\frac{1}{2}$
12	-2	$A +\frac{1}{2}$	$bq_1 -\frac{1}{2}$	$C +\frac{1}{2}$	$dq_2 -\frac{1}{2}$	$e -\frac{1}{2}$
13	0	$A +\frac{1}{2}$	$bq_1 -\frac{1}{2}$	$c -\frac{1}{2}$	$DQ_2 +\frac{1}{2}$	$E +\frac{1}{2}$
14	0	$A +\frac{1}{2}$	$bq_1 -\frac{1}{2}$	$c -\frac{1}{2}$	$DQ_2 +\frac{1}{2}$	$e -\frac{1}{2}$
15	-2	$A +\frac{1}{2}$	$bq_1 -\frac{1}{2}$	$c -\frac{1}{2}$	$dq_2 -\frac{1}{2}$	$E +\frac{1}{2}$
16	-2	$A +\frac{1}{2}$	$bq_1 -\frac{1}{2}$	$c -\frac{1}{2}$	$dq_2 -\frac{1}{2}$	$e -\frac{1}{2}$
17	+2	$a -\frac{1}{2}$	$BQ_1 +\frac{1}{2}$	$C +\frac{1}{2}$	$DQ_2 +\frac{1}{2}$	$E +\frac{1}{2}$
18	+2	$a -\frac{1}{2}$	$BQ_1 +\frac{1}{2}$	$C +\frac{1}{2}$	$DQ_2 +\frac{1}{2}$	$e -\frac{1}{2}$
19	0	$a -\frac{1}{2}$	$BQ_1 +\frac{1}{2}$	$C +\frac{1}{2}$	$dq_2 -\frac{1}{2}$	$E +\frac{1}{2}$
20	0	$a -\frac{1}{2}$	$BQ_1 +\frac{1}{2}$	$C +\frac{1}{2}$	$dq_2 -\frac{1}{2}$	$e -\frac{1}{2}$
21	+2	$a -\frac{1}{2}$	$BQ_1 +\frac{1}{2}$	$c -\frac{1}{2}$	$DQ_2 +\frac{1}{2}$	$E +\frac{1}{2}$
22	+2	$a -\frac{1}{2}$	$BQ_1 +\frac{1}{2}$	$c -\frac{1}{2}$	$DQ_2 +\frac{1}{2}$	$e -\frac{1}{2}$
23	0	$a -\frac{1}{2}$	$BQ_1 +\frac{1}{2}$	$c -\frac{1}{2}$	$dq_2 -\frac{1}{2}$	$E +\frac{1}{2}$
24	0	$a -\frac{1}{2}$	$BQ_1 +\frac{1}{2}$	$c -\frac{1}{2}$	$dq_2 -\frac{1}{2}$	$e -\frac{1}{2}$
25	0	$a -\frac{1}{2}$	$bq_1 -\frac{1}{2}$	$C +\frac{1}{2}$	$DQ_2 +\frac{1}{2}$	$E +\frac{1}{2}$
26	0	$a -\frac{1}{2}$	$bq_1 -\frac{1}{2}$	$C +\frac{1}{2}$	$DQ_2 +\frac{1}{2}$	$e -\frac{1}{2}$
27	-2	$a -\frac{1}{2}$	$bq_1 -\frac{1}{2}$	$C +\frac{1}{2}$	$dq_2 -\frac{1}{2}$	$E +\frac{1}{2}$
28	-2	$a -\frac{1}{2}$	$bq_1 -\frac{1}{2}$	$C +\frac{1}{2}$	$dq_2 -\frac{1}{2}$	$e -\frac{1}{2}$

(continued)

**Box 6.5** (continued)

Line	Yield mean 8	Locus A	Locus B and $Q1$	Locus C	Locus D and $Q2$	Locus E
29	0	$a - \frac{1}{2}$	$bq_1 - \frac{1}{2}$	$c - \frac{1}{2}$	$DQ_2 + \frac{1}{2}$	$E + \frac{1}{2}$
30	0	$a - \frac{1}{2}$	$bq_1 - \frac{1}{2}$	$c - \frac{1}{2}$	$DQ_2 + \frac{1}{2}$	$e - \frac{1}{2}$
31	-2	$a - \frac{1}{2}$	$bq_1 - \frac{1}{2}$	$c - \frac{1}{2}$	$dq_2 - \frac{1}{2}$	$E + \frac{1}{2}$
32	-2	$a - \frac{1}{2}$	$bq_1 - \frac{1}{2}$	$c - \frac{1}{2}$	$dq_2 - \frac{1}{2}$	$e - \frac{1}{2}$

$$\hat{Y} = b_0 + b_A X_A + b_B X_B + \dots + b_E X_E$$

where  $b_0$  is intercept ( $b_0 = \mu_{(Y)} - \sum b \mu_{(X)}$ ),  $b_A, b_B \dots b_E$  are the regression coefficients for predictors  $X_A, X_B \dots X_E$ , respectively (taking values of 1 and 0 for  $A$  and  $a, B$  and  $b$ , etc.), and are estimated to minimize  $\sum(Y - \hat{Y})^2$  (least squares estimate),  $Y$  being the observed yield and  $\hat{Y}$  the estimated yield. We will in fact work with departures from the means for the  $Y$  values ( $y = Y - \mu_{(Y)}$ ) and the  $X$ 's (means of  $X$ 's are all  $\frac{1}{2}$  so values of  $x$ 's are  $+\frac{1}{2}$  or  $-\frac{1}{2}$ ):

$\hat{y} = b_A x_A + b_B x_B + \dots + b_E x_E$ , which in matrix algebra is:

$\hat{y} = \mathbf{X}\mathbf{b}$  where  $\mathbf{y}$  (32 rows) and  $\mathbf{b}$  (5 rows) are column vectors and  $\mathbf{X}$  is a matrix (32 rows and 5 columns).

The intercorrelation matrix for the pairs of columns ( $\mathbf{y}, \mathbf{x}_A, \mathbf{x}_B, \mathbf{x}_C, \mathbf{x}_D$  and  $\mathbf{x}_E$ ) is:

$\mathbf{R} =$	$y$	$x_A$	$x_B$	$x_C$	$x_D$	$x_E$
$y$	1.0	0.0	0.7071	0.0	0.7071	0.0
$x_A$	0.0	1.0	0.0	0.0	0.0	0.0
$x_B$	0.7071	0.0	1.0	0.0	0.0	0.0
$x_C$	0.0	0.0	0.0	1.0	0.0	0.0
$x_D$	0.7071	0.0	0.0	0.0	1.0	0.0
$x_E$	0.0	0.0	0.0	0.0	0.0	1.0

We therefore have a very simple intercorrelation matrix in which yield is correlated with presence and absence of marker at loci  $B$  and  $D$  ( $r = 0.7071$ ) but not at  $A, C$  and  $E$  ( $r = 0$ ), and there are no correlations between markers (all 0).

The least squares estimates of  $\mathbf{b}$  (which equal the maximum likelihood estimates for regression models with normal errors) are:

$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$  where  $(\mathbf{X}'\mathbf{X})^{-1}$  is the inverse of  $(\mathbf{X}'\mathbf{X})$  and  $\mathbf{X}'$  is the transpose of  $\mathbf{X}$ .

The derivation can be found in any standard textbook on statistics.

In our simple example:

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{vmatrix} \frac{1}{8} & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{8} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{8} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{8} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{8} \end{vmatrix} \quad \mathbf{X}'\mathbf{y} = |0| \quad \mathbf{b} = |0|$$

$$\begin{array}{ccc} |16| & & |2| \\ |0| & & |0| \\ |16| & & |2| \\ |0| & & |0| \end{array}$$

In other words,  $b_A = b_C = b_E = 0.0$ , and  $b_B = b_D = 2.0$ ; and the predicted yields ( $\hat{y}$ ) equal the actual yields ( $y$ ). The  $b$ 's are the (genetic) effects of changing allele  $a$  to  $A$ ,  $b$  to  $B$ ,  $c$  to  $C$ ,  $d$  to  $D$  and  $e$  to  $E$ , respectively.

The total sum of squares for yield is  $\mathbf{y}'\mathbf{y} = 64$

The sum of squares due to regression is  $\mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b} = 64$ , and hence accounts for all of the variation. As  $\mathbf{b}'\mathbf{X}'\mathbf{X} = [0 \ 16 \ 0 \ 16 \ 0]$  and  $\mathbf{b}' = [0 \ 2 \ 0 \ 2 \ 0]$ , each marker accounts for half of this variation (correlation between yield and each marker was  $r = 0.7071$ ; hence  $r^2 = 0.5$  for each marker).

Furthermore  $R^2 = \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b}/\mathbf{y}'\mathbf{y} = 1.0$  in our simple example, where  $R$  is the multiple correlation coefficient  $R$  (the correlation between  $y$  and  $\hat{y}$ ).

In practice there will be environmental effects ( $\varepsilon$ 's) as well as genetic effects so the model for  $n$  genotypes ( $i = 1 \dots n$ ) and  $p$  markers ( $j = 1 \dots p$ ) will be:

$$y_i = g_i + \varepsilon_i = \sum b_j x_{ij} + \varepsilon_i, \text{ where summation is over the } p \text{ markers.}$$

The predicted yields ( $\hat{y} = \sum b_j x_{ij}$ ) will no longer equal the actual yields ( $y$ ).

With clones rather than inbred lines, the  $x_i$  may have three genotypes  $AA$ ,  $Aa$  and  $aa$ , which can be coded 2, 1 and 0 for an additive model; and the genetic model can be extended to accommodate dominance and epistasis between loci. It should also be pointed out that allele frequencies can vary across loci so that this can affect their contribution to the genetic variance as well as the size of their effects.

## Issues in Practice for GWAS and GS

A number of issues need to be addressed in the practice of genome-wide association genetics (GWAS) and genomic selection (GS). They usually involve scoring tens-of-thousands of markers, thus creating a computational problem which needs to be solved. In multiple regression analysis, the number ( $n$ ) of genotypes is expected to be larger than the number of markers ( $p$ ) whereas we are now in a situation where the latter is going to be considerably larger than the former (Meuwissen et al. 2001). Fitting this large- $p$  with small- $n$  regression requires using some type of variable selection or shrinkage estimation procedure, details of which can be found in the review by de los Campos et al. (2013). Shrinkage estimation avoids having to select markers for inclusion in model.

## ***Linkage Disequilibrium***

The effectiveness of GWAS depends on the extent and magnitude of linkage disequilibrium, which needs to be determined in the germplasm under consideration. Ideally one would like to find that markers very close together are in linkage equilibrium (no association) so that a QTL is associated with one marker, but not its neighbouring markers. In rice (*Oryza sativa*), Mather et al. (2007) found the extent of linkage disequilibrium to be greatest in temperate *japonica* (>500 kb), followed by tropical *japonica* (about 150 kb) and *indica* (about 75 kb). The distance was shorter in their wild ancestor, *O. rufipogon* (<40 kb), presumably because it outcrosses at a much higher rate (7–56 %) than domesticated rice (1–2 %). The authors point out that 75 and 150 kb correspond to genomic regions encompassing 9 and 17 genes, respectively. Furthermore, given the genome size of rice, placement of SNP markers every 75 kb in *indica* rice would require a total of 5200 for genome wide coverage, whereas every 150 kb in tropical *japonica* would require 2600 markers. Hence association mapping is feasible with a more modest number of SNPs than indicated earlier. The authors compare their results with those reported in the literature for other species, namely 2 kb in outcrossing maize, 20 kb in sorghum (outcrossing 10–20 %), and several hundred kilobases in domesticated barley and soybean (inbreeding species). For any particular crop, however, a breeder should check the literature for reports on the extent of linkage disequilibrium.

## ***Population Structure***

One needs to be sure that factors other than very tight linkage are not the cause of the linkage disequilibrium in the germplasm; for example, population structure. Let us consider a simple but artificial example: 32 inbred lines, two unlinked markers *A* and *B* in linkage equilibrium and a QTL for yield ( $QQ = 9$  t/ha and  $qq = 7$  t/ha) which is also unlinked to *A* and *B* (Box 6.6). As in the previous example, we determine the yields (*Y*) of the 32 (*N*) inbred lines and their marker allele composition and do a standard multiple regression analysis.

$\hat{y} = b_A x_A + b_B x_B$ , which in matrix algebra is:

$\hat{y} = \mathbf{X}\mathbf{b}$  where  $\mathbf{y}$  (32 rows) and  $\mathbf{b}$  (2 rows) are column vectors and  $\mathbf{X}$  is a matrix (32 rows and 2 columns).

The intercorrelation matrix for the pairs of columns ( $\mathbf{y}$ ,  $\mathbf{x}_A$ ,  $\mathbf{x}_B$ ) is:

$$\mathbf{R} = \begin{array}{|ccc|} \hline & y & x_A & x_B \\ \hline y & 1.0 & 0.25 & 0.0 \\ x_A & 0.25 & 1.0 & 0.0 \\ x_B & 0.0 & 0.0 & 1.0 \\ \hline \end{array}$$

We therefore have a very simple intercorrelation matrix in which yield is correlated with presence and absence of marker at locus  $A$  ( $r=0.25$ ) but not at  $B$  ( $r=0$ ), and there is no correlation between marker  $A$  and  $B$  ( $r=0$ ).

$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$  where  $(\mathbf{X}'\mathbf{X})^{-1}$  is the inverse of  $(\mathbf{X}'\mathbf{X})$  and  $\mathbf{X}'$  is the transpose of  $\mathbf{X}$ .

In our simple example:

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{vmatrix} \frac{1}{8} & 0 \\ 0 & \frac{1}{8} \end{vmatrix} \quad \mathbf{X}'\mathbf{y} = \begin{vmatrix} 4 \\ 0 \end{vmatrix} \quad \mathbf{b} = \begin{vmatrix} 0.5 \\ 0.0 \end{vmatrix}$$

In other words,  $b_A = 0.5$  and  $b_B = 0.0$ ; and the predicted yields ( $\hat{y}$ ) equal one quarter of the actual yields ( $y$ ). The total sum of squares for yield is  $\mathbf{y}'\mathbf{y} = 32$ .

The sum of squares due to regression is  $\mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b} = 2$ , and hence accounts for 1/16th of the variation. As  $\mathbf{b}'\mathbf{X}'\mathbf{X} = [4 \ 0]$  and  $\mathbf{b}' = [0.5 \ 0]$ , marker  $A$  accounts for this variation (correlation between yield and marker  $A$  was  $r=0.25$ ; hence  $r^2=1/16$ ).

### Box 6.6

Association genetics: yields of 32 inbred lines (departures from mean of 8) and marker genotypes at two unlinked loci  $A$  and  $B$  in linkage equilibrium; there is one QTL ( $Q$ ) which is also unlinked to  $A$  and  $B$

Line	Yield mean 8	Locus $A$	Locus $Q$	Locus $B$
1	+1	$A +\frac{1}{2}$	$Q$	$B +\frac{1}{2}$
2	+1	$A +\frac{1}{2}$	$Q$	$B +\frac{1}{2}$
3	+1	$A +\frac{1}{2}$	$Q$	$B +\frac{1}{2}$
4	+1	$A +\frac{1}{2}$	$Q$	$B +\frac{1}{2}$
5	+1	$A +\frac{1}{2}$	$Q$	$B +\frac{1}{2}$
6	+1	$A +\frac{1}{2}$	$Q$	$b -\frac{1}{2}$
7	+1	$A +\frac{1}{2}$	$Q$	$b -\frac{1}{2}$
8	+1	$A +\frac{1}{2}$	$Q$	$b -\frac{1}{2}$
9	+1	$A +\frac{1}{2}$	$Q$	$b -\frac{1}{2}$

(continued)

**Box 6.6** (continued)

Line	Yield mean 8	Locus A	Locus Q	Locus B
10	+1	$A +\frac{1}{2}$	$Q$	$b -\frac{1}{2}$
11	-1	$A +\frac{1}{2}$	$q$	$B +\frac{1}{2}$
12	-1	$A +\frac{1}{2}$	$q$	$B +\frac{1}{2}$
13	-1	$A +\frac{1}{2}$	$q$	$B +\frac{1}{2}$
14	-1	$A +\frac{1}{2}$	$q$	$b -\frac{1}{2}$
15	-1	$A +\frac{1}{2}$	$q$	$b -\frac{1}{2}$
16	-1	$A +\frac{1}{2}$	$q$	$b -\frac{1}{2}$
17	-1	$a -\frac{1}{2}$	$q$	$B +\frac{1}{2}$
18	-1	$a -\frac{1}{2}$	$q$	$B +\frac{1}{2}$
19	-1	$a -\frac{1}{2}$	$q$	$B +\frac{1}{2}$
20	-1	$a -\frac{1}{2}$	$q$	$B +\frac{1}{2}$
21	-1	$a -\frac{1}{2}$	$q$	$B +\frac{1}{2}$
22	-1	$a -\frac{1}{2}$	$q$	$b -\frac{1}{2}$
23	-1	$a -\frac{1}{2}$	$q$	$b -\frac{1}{2}$
24	-1	$a -\frac{1}{2}$	$q$	$b -\frac{1}{2}$
25	-1	$a -\frac{1}{2}$	$q$	$b -\frac{1}{2}$
26	-1	$a -\frac{1}{2}$	$q$	$b -\frac{1}{2}$
27	+1	$a -\frac{1}{2}$	$Q$	$B +\frac{1}{2}$
28	+1	$a -\frac{1}{2}$	$Q$	$B +\frac{1}{2}$
29	+1	$a -\frac{1}{2}$	$Q$	$B +\frac{1}{2}$
30	+1	$a -\frac{1}{2}$	$Q$	$b -\frac{1}{2}$
31	+1	$a -\frac{1}{2}$	$Q$	$b -\frac{1}{2}$
32	+1	$a -\frac{1}{2}$	$Q$	$b -\frac{1}{2}$

In summary, despite the fact that  $A$ ,  $B$  and  $Q$  are unlinked, the first marker does account for some of the variation in yield, albeit only 1/16th. The explanation lies in the way I constructed the population of inbred lines. The population is in fact an equal mixture of two populations each in linkage equilibrium. The frequencies ( $P$ ) of  $AQ$ ,  $Aq$ ,  $aQ$  and  $aq$  in the first population are 9/16, 1/16, 3/16 and 3/16 whereas in the second population they are 1/16, 9/16, 3/16 and 3/16. Hence in the mixed population the frequencies are 5/16, 5/16, 3/16 and 3/16. The amount of linkage disequilibrium is  $D = P(AQ)P(aq) - P(Aq)P(aQ) = 1/16$  ( $D = 0$  when there is no disequilibrium: see Chap. 9 for more details). If the 32 inbred lines were to be crossed in all combinations and a new set of inbred lines (doubled haploids) produced from the resulting hybrids, the amount of linkage disequilibrium ( $D$ ) would be halved to 1/32; and in each further cycle by half again. In other words, marker  $A$  would become a less and less effective predictor of yield; it is in effect a false positive association of marker and trait.

The problem of population structure arises because any phenotypic trait that is also correlated with the underlying population structure at marker loci will show an inflated number of (false) positive associations. Details of population structure can

be inferred from a set of unlinked genetic markers and used to identify homogeneous subpopulations within which mating is assumed random (Pritchard et al. 2000); although the software package STRUCTURE has been applied to collections of clones and inbred lines. In other words, Pritchard et al. (2000) used multilocus genotype data to infer population structure and assign individuals to populations. They assumed a model in which there are  $K$  populations (where  $K$  may be unknown), each of which is characterized by a set of allele frequencies at each locus. Individuals in the sample are assigned (probabilistically) to populations, or jointly to two or more populations if their genotypes indicate that they are admixed. Their model accounts for the presence of Hardy-Weinberg or linkage disequilibrium by introducing population structure that attempts to find subpopulation groupings that are not in disequilibrium. All of the available markers are then tested for associations within these sub-populations. Later STRUCTURE was extended to linked markers by Falush et al. (2003). Price et al. (2006) proposed principal component analysis as an alternative method to correct for stratification in genome-wide association studies.

## ***Kinship***

Kinship is another potential problem. An extreme case would be a collection of inbred lines containing more than one representative of some lines. For two unlinked loci in equilibrium, a random sample of lines might have equal numbers of genotypes  $AABB$ ,  $AAbb$ ,  $aaBB$  and  $aabb$ , but if additional representatives of the first and last genotypes were present, the sample might be  $AABB$ ,  $AABB$ ,  $AABB$ ,  $AAbb$ ,  $aaBB$ ,  $aabb$ ,  $aabb$  and  $aabb$ . The frequencies of gametes produced from these genotypes would be  $P(AB) = \frac{3}{8}$ ,  $P(Ab) = \frac{1}{8}$ ,  $P(aB) = \frac{1}{8}$  and  $P(ab) = \frac{3}{8}$ , so that the estimated level of disequilibrium would be  $D = \frac{1}{8}$ . In actual collections of germplasm one could have parents and offspring, full-sibs, half-sibs, cousins and other relatives; and such kinship could create linkage disequilibrium.

## ***Population Structure and Kinship***

Yu et al. (2005) showed how information on population structure and kinship can be incorporated into the traditional mixed-model used by animal breeders (see earlier section on BLUP), through matrices usually labelled **Q** (obtained using STRUCTURE) and **K**, so that spurious associations do not arise from these sources. These authors demonstrated the superiority of the complete **QK** model over simpler models for flowering time, ear height and ear diameter in 277 diverse maize inbred lines. The **K** matrix, as well as the **Q** matrix, can be derived from the marker data. Coancestry coefficients calculated from pedigree records are not required but it is

necessary to define what is meant by unrelated individuals (random pairs of inbred lines are one arbitrary definition, another is in terms of unshared alleles). Stich et al. (2008) confirmed through empirical data and simulations that the **QK** method is also appropriate for autogamous species such as wheat. They examined the grain yield of 303 soft winter wheat inbreds and determined **Q** and **K** matrices using 37 (36 SSR) marker loci randomly distributed across 19 out of wheat's 21 chromosomes. They did, however, recommend replacing the **K** matrix by a **K<sub>T</sub>** one based on a REML estimate of the probability that two inbreds carry alleles at the same locus that are identical in state but not identical by descent. In barley, Cockram et al. (2010) used **K** alone to achieve successful genome wide association mapping in the presence of extremely high population substructure. More recently, Würschum and Kraft (2015) showed that multi-locus models are an attractive alternative for genome-wide association studies in which markers are selected as cofactors to control population structure and genetic background variation. They applied the method to their sugar beet data which is presented in the section below on cross-validation. Clearly choosing the best method for a GWAS analysis is an active area of research.

## ***GS Methods***

A number of GS methods have been suggested and their accuracy tested, including the use of ridge regression (Hoerl and Kennard 1970) (RR-BLUP), the Least Absolute Shrinkage and Selection Operator (LASSO) (Tibshirani 1996), their Bayesian counterparts and others mentioned by Crossa et al. (2010). The methods differ in the extent to which they take account of linkage disequilibrium, population structure and kinship and hence can give different results. Jonas and de Koning (2013) concluded from their review that further method development for estimating genomic breeding values in crop species is required to capture nonadditive effects, G × E, and crop-specific (in)breeding cycles.

Unlike fixed linear regression models, linear mixed models have no well-established  $R^2$  statistic for assessing goodness-of-fit and prediction power. Sun et al. (2010) showed that the likelihood-ratio-based  $R^2$  ( $R^2_{LR}$ ) satisfies several desirable requirements for a test statistic. As it reduces to the regular  $R^2$  for fixed models without random effects other than residual, it provides a general measure for the effect of QTL in mixed-model association mapping.

## Cross-Validation

Cross-validation has proved a useful approach for obtaining unbiased estimates of QTL effects and determining the magnitude of bias of predictive power in GWAS and GS. QTL detection and effect estimation are done in a subset of the lines (estimation set, ES) while the remaining lines of the population are set aside to serve as an independent validation set (test set, TS). Cross-validation mimics what happens in plant breeding where the breeder wants to use QTL detected in one population in different but related germplasm. Würschum et al. (2014) used a large sugar beet (*Beta vulgaris*) data set to evaluate the potential of cross-validation with the following percentages of TS: 5, 10, 20, 30, 40 and 50 %. They produced testcross progenies by crossing 924 diploid elite sugar beet inbred lines to a single-cross hybrid as tester. The 924 genotypes were fingerprinted with 677 SNP markers. The association mapping used a mixed model in which the allele substitution effect was modelled as a fixed effect and the genetic effects of the lines as a random effect. The variance of the random genetic effects was estimated by REML and multiplied by a  $924 \times 924$  matrix of kinship coefficients (calculated from marker data) that defined the degree of genetic covariance between all pairs of entries. Würschum et al. (2014) concluded that the optimum proportion of plants to be used as TS is dependent on the population size, but for commonly used population sizes (a few hundred individuals), fivefold cross-validation (TS = 20 %) is appropriate. Applying fivefold cross-validation in the full data set, they observed for all six traits assessed (range of 4–15 QTLs), a substantially reduced cross-validated proportion of explained genotypic variance compared with that estimated in the full data set (range for six traits was 73 to 10 %). For all traits the relative bias averaged 38 %. Accurate assessments of the proportion of genotypic variance explained by the QTLs is important in deciding whether or not marker-assisted selection for these QTLs is worthwhile in a breeding programme, compared with phenotypic selection. Finally, most of the QTLs detected with the full data set were identified in at least 40 % of the 1000 fivefold cross-validation runs.

## GS Results in Practice

Since 2010 an increasing number of papers have appeared in the literature and more can be anticipated; for example, in forest trees such as Interior Spruce (*Picea engelmannii* × *glauca*) (Ratcliffe et al. 2015). Here just a few results for wheat and maize will be mentioned as examples of current research interests. Crossa et al. (2010) reported results from two distinct data sets. The first one was from a collection of 599 historical CIMMYT wheat breeding lines, assessed for grain yield in four target sets of environments. Pedigree information was available for the lines which were also genotyped using 1447 Diversity Array Technology markers. The

second data set was from 300 CIMMYT maize lines from their Drought Tolerance Maize for Africa project which had been genotyped with 1148 SNP markers. Grain yield, days to silking and days to anthesis were assessed under severe drought stress and well-watered conditions. Crossa et al. (2010) evaluated several models for GS that differed in the type of information used for constructing predictions (pedigree, markers, or both) and in how molecular markers were incorporated into the model (parametric versus semiparametric methods). Separate models were fitted to each trait–environment combination. In other words, the phenotypic value of line  $i$  was:

$$y_i = \mu + g_i + e_i$$

where  $y_i$  is the average performance of line  $i$  for the trait–environment combination,  $\mu$  is the ‘intercept’ for the trait–environment combination,  $g_i$  is the genetic value of line  $i$  and  $e_i$  is the residual in the model. Models differed in how pedigree and molecular marker information was included in  $g_i$ . The theory underlying them and estimation procedures are difficult for non statisticians, and hence plant breeders will certainly require help from statisticians in their use and interpretation. Here we will simply look at the conclusions of Crossa et al. (2010), particularly those from cross-validation which assessed the ability of a model to predict future outcomes by using estimates from 90 % of data to predict the phenotypes of the other 10 %. In wheat for grain yield the correlations between observed and predicted values ranged from 0.355 to 0.608. Models jointly using marker and pedigree data outperformed both pedigree and marker models across traits and environments, regardless of the choice of model. In maize, for flowering time the correlations varied from 0.464 to 0.79 and for grain yield from 0.415 to 0.514. The authors did not find a method (parametric versus semiparametric) that was consistently superior across environments and traits; but it is perhaps too early to say that this is going to be a general finding in plant breeding. Finally they concluded that there is room for improving predictive ability even further, based on the maximum attainable correlation. This occurs when the cross-validation predictions actually equal the genetic values ( $g_i$ 's). Then

$$y_i = \mu + g_i + e_i$$

and the correlation between  $y_i$  and  $g_i$  is

$$\begin{aligned} \text{Cov}(y_i, g_i) / (V y_i V g_i)^{\frac{1}{2}} &= \\ V g_i / [(V g_i + V e_i) V g_i]^{\frac{1}{2}} &= \\ (V g_i)^{\frac{1}{2}} / (V g_i + V e_i)^{\frac{1}{2}} &= \\ [V g_i / (V g_i + V e_i)]^{\frac{1}{2}} &= h, \text{ the square root of the heritability.} \end{aligned}$$

Hence if the heritability is 0.5, the maximum correlation is 0.707. However, with  $r$  replicates, the maximum correlation is higher as

$$h(r) = [Vg_i/(Vg_i + Ve_i/r)]^{1/2}$$

If the correlation between predicted and true genotypic values is actually  $r_{MG}$ , then the correlation between the predicted and observed values is  $r_{MP} = hr_{MG}$

The highest correlations found by Crossa et al. (2010) ranged from 0.40 to 0.79 and were well below the theoretical maxima given by the heritabilities of the traits and the number of replicates. Better predictive ability should be possible when more markers or better models are available.

The above results came from analyses of germplasm collections. The next example is perhaps more typical of the use of GS in breeding programmes.

### ***Genomic Prediction in Tropical Maize Breeding***

Zhang et al. (2015) compared genomic prediction accuracy using low density markers (162–212 SNPs per population) with those of high density genotyping-by-sequencing (GBS, 48,662–78,005 SNPs per population) in 19 bi-parental populations of tropical maize, derived from 23 elite maize inbred lines. Between 126 and 184 lines from the F<sub>2</sub> (or BC<sub>1</sub>F<sub>2</sub>) populations were crossed to an F<sub>1</sub> tester and the progenies evaluated in multi-environment trials in both water-stressed and well-watered environments in Africa. Genotypic values were predicted using an extension of the Genomic Best Linear Unbiased Predictor (GBLUP) that incorporates genotype × environment (GE) interaction (i.e. correlated environmental structures). For cross-validation, 80 % of each population was used for ‘training’ and 20 % for ‘testing/validation’, with the five permutations of five subsets repeated 20 times to give a total of 100 runs. Prediction accuracy ( $r_{MG}$ , see above) averaged over the 19 populations ranged from 0.25 to 0.45. The results showed that low density markers were largely sufficient for good prediction in bi-parental maize populations for simple traits (anthesis date and plant height) with moderate-to-high heritability, but GBS outperformed low density SNPs for complex traits (grain yield) and simple traits evaluated under stress conditions with low-to-moderate heritability. Prediction accuracy of complex traits was consistently lower than that of simple traits, and prediction accuracy under stress conditions was consistently lower and more variable than under well-watered conditions for all the target traits due to their poor heritability under stress conditions. Models incorporating GE were superior for complex traits but only marginally better for simple traits.

## Implications for Breeders

The current situation for potato breeders is typical of that for many crops (Bradshaw and Bonierbale 2010). Major genes have been mapped for flesh, skin and flower colour, for tuber shape and eye depth, and for resistances to late blight, cyst nematodes, root-knot nematodes, viruses (PVY, PVA, PVX, PVM, PVS and PLRV), and wart. Large-effect QTLs have been mapped for total glycoalkaloid content (TGA), maturity and resistances to late blight, *Verticillium* wilt, cyst nematodes and PLRV. Furthermore, diagnostic markers for these major genes and QTL alleles of large effect are becoming available for marker-assisted selection. However, many economically important traits are still best viewed as complex polygenic traits, despite a number of QTLs being found, and these include dormancy, dry matter and starch content, fry colour, resistance to *Pectobacterium* (*Erwinia*), tuberization and yield. For these traits breeders still rely primarily on phenotypic data and the concepts of quantitative genetics to determine crossing and selection strategies, but there is increasing interest in the possibility of genomic selection following its implementation in cattle (dairy) breeding (Jonas and de Koning 2013). Crossa et al. (2014) have described and summarized the results of genomic prediction in CIMMYT's maize and wheat breeding programmes starting from an initial assessment of the predictive ability of different models using pedigree and marker information to the present, when methods for implementing GS in practical global maize and wheat breeding programmes are being studied and investigated. Results of analyzing vast amounts of data in both CIMMYT breeding programmes indicate that pedigree and markers offer opportunities for achieving prediction that can be exploited in the breeding pipeline. However, questions on where and how to use this information remain open.

The implications for breeders are clear: for as long as some phenotypic selection is required for quantitative traits in a breeding programme, at least some generations of the programme will need to be assessed in field trials. Furthermore, the programme design will need to accommodate any genotype  $\times$  environment interactions, the topic considered in the next chapter. The challenge for breeders is how best to prioritize and integrate selection of desired alleles for qualitative traits and ones of large effect at QTLs with field selection for quantitative traits; having first ensured that they are present among the parental germplasm. A new challenge is if and when to integrate genomic selection. All plant breeding comprises cycles of crossing and selection (recurrent selection). Hence it could be argued that the inbred lines (and their hybrids) and clones being evaluated in one cycle should be the training population which is phenotyped as well as genotyped, and then the next generation of lines or clones would be the selection population that is initially only genotyped. In other words, these training and selection populations should be closely related enough for genomic selection to be effective.

# **Chapter 7**

## **Genotype x Environment Interactions and Selection Environments**

### **Introduction**

A key strategic decision at the start of any breeding programme is the number of new cultivars required for a given range of target environments and different end uses. It is unlikely that one of many potential new cultivars will be best in all environments and for all uses. Often crops can be divided into distinct varietal groups in terms of both agroecological adaptations and end-use properties, each group requiring its own breeding programme. Take three wheat examples. Firstly, winter wheat in the USA can be divided into four groups: (1) Eastern and South-eastern soft wheats; (2) Southern Great Plains wheats; (3) Northern Great Plains wheats; and (4) Pacific Northwest wheats (Baenziger et al. 2009). Secondly, spring wheat in the USA can be divided into three groups based on kernel colour, endosperm hardness and other quality characteristics: (1) hard red spring wheat; (2) soft white wheat; and (3) hard white spring wheat (Mergoum et al. 2009a). Of these groups, it is hard red spring wheat that contains the highest percentage of protein and has the best milling and baking characteristics, making it ideal for bread. Thirdly, in CIMMYT's spring wheat programme, co-ordinated from Mexico, breeding objectives and the assessment of germplasm are targeted at five mega-environments (not ME3) which are geographical areas where wheat adaptation can be expected to be similar in terms of climatic, disease, or crop-management constraints (Singh and Trethowan 2007): ME1 is irrigated, moderate temperature; ME2 high rainfall >500 mm, temperate; ME3 high rainfall, acid soil; ME4 low rainfall <500 mm, temperate or hot; ME5 irrigated or high rainfall, warmer areas between 23°N and 23°S and below 1000 m altitude; and ME6 low rainfall, cool areas >45°N or S.

When a crop like a major food staple is grown in different parts of the world with different lengths of growing season one might anticipate the need for a number of cultivars with different maturities to cover the range, despite any desire to breed for general adaptability. Likewise, if a crop is daylength dependent *per se*, cultivars

adapted to a certain latitude cannot be grown commercially in other latitudes, a good example being onions (Shigyo and Kik 2008). In contrast, there is little evidence for a photoperiod effect on carrot root production and flowering, but adaptation to a wide range of production temperatures is an issue (Simon et al. 2008). Climate will determine whether to breed cultivars of many crops for winter (autumn) or spring sowing. Successful general purpose cultivars can be found, but one would expect different barley cultivars for use as animal feed and for malting, and different potato cultivars for cooking as a vegetable (baking and boiling) and processing into products (French fries and crisps). In vegetables such as onions, different cultivars are required for the storage and fresh markets which can be equally important. Another example comes from soybean, where processors generally consider that Chinese and Japanese soybeans are better suited for production of traditional foods because they are higher in protein, whereas North American soybeans appear more suitable for oil extraction (Vollman and Menken 2012). This view is supported by studies of long-term trends in North American soybean breeding where progress for higher yield was frequently associated with reduced protein content, whereas protein content of Chinese cultivars remained constant (Vollman and Menken 2012).

In some situations it is not obvious how many different kinds of cultivars are required. For example, can cultivars be bred with tolerance of abiotic stresses but which also yield well in the absence of such stresses, or are two different types of cultivar required? Here tolerance means little reduction of yield under stress compared with optimum growing conditions. Likewise, can cultivars be bred with high resource use efficiency under both high and low inputs of the resource? For example, can cultivars be bred for higher levels of production than other cultivars on organic farms, low-input farms and high-input farms, or are different cultivars needed for each farming system? Finally, can cultivars be bred that do well under both zero-tillage and conventional tillage farming? Ultimately, an understanding of the physiological and biochemical bases of stress tolerance and resource use efficiency, and their implications for yield potential, are required if the breeder is to be provided with selection criteria for manipulating G × E.

Soybean provides a different kind of example where an assessment of genotype × environment interactions is required. Luna and Planchon (1995) reported a soybean genotype (4 genotypes) by rhizobial strain (2 strains) (*Bradyrhizobium japonicum*) interaction for symbiotic nitrogen fixation activity. Thus the replacement of one strain with another might result in the alteration of the relative agronomic performances of the soybean cultivars since nitrogen fixation is considered a major factor of soybean productivity.

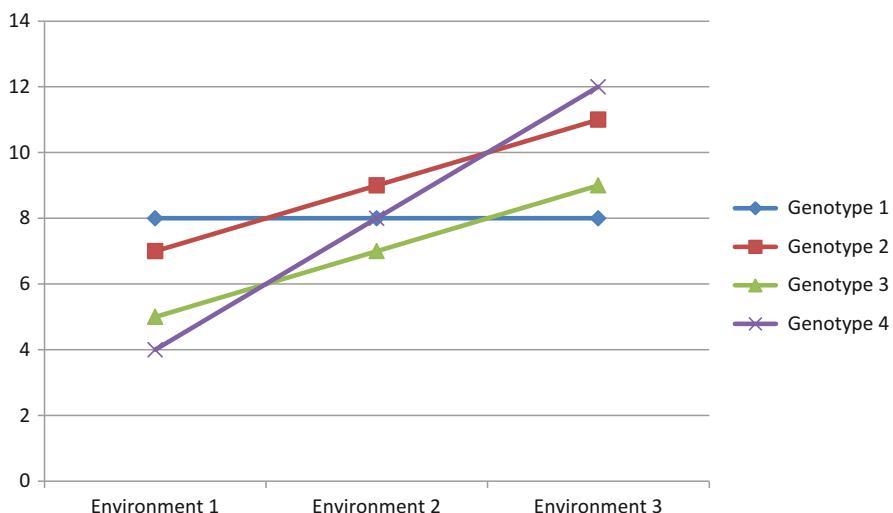
An assessment of genotype × environment interactions is required at the start of breeding programmes in order to determine objectives and to choose parents for hybridization, and at the end of programmes in order to make recommendations to farmers and growers about which cultivars to grow. Equally important for the success of a breeding programme is the choice of environment at every intermediate stage where families, inbred lines, hybrids or clones are assessed and selected, depending on the type of cultivar being sought. Furthermore, all breeders face the

issue of the effectiveness of selecting individual plants grown at wide spacing compared with assessing potential cultivars under normal growing density; for example, spaced plants versus swards in grass breeding (Posselt 2010). This chapter explains the concept of genotype  $\times$  environment interactions, how they are assessed and their consequences for plant breeding.

## Concept of Genotype $\times$ Environment Interactions ( $G \times E$ )

Genotype  $\times$  environment interactions can occur between genotypes and specific macroenvironments such as locations, seasons and fertilizer regimes. The interaction may or may not involve changes in the ranking of genotypes over environments. The literature on genotype  $\times$  environment interactions is vast and the methods of analysis are numerous, but there are relatively few underlying concepts. Even so, the topic continues to attract controversy and debate, as seen by two reviews in *Crop Science* (Yan et al. 2007; Gauch et al. 2008). A simple (artificial) example is shown in Fig. 7.1 for the yields (phenotypes  $P$ ) of four genotypes in three environments.

Genotype 1 is the best one in environment 1, genotype 2 in environment 2, and genotype 4 in environment 3. Hence the results of any genetic analysis will depend upon the environment in which the genotypes are assessed. Likewise, if a breeding programme was conducted in low-yielding environment 1, genotype 1 would be selected but would be the worst if grown in environment 3. In contrast, genotype 4 would be selected in high-yielding environment 3 but would be the worst if grown in environment 1. Interestingly, genotype 2 which would be selected in environment



**Fig. 7.1** Yields of four genotypes grown in three environments

2 would rank second in both environments 1 and 3. It would also be selected on the basis of the average of yields in environments 1 and 3. Genotype 2 is high yielding with general adaptability whereas genotypes 1 and 4 have specific adaptability to low and high yielding environments. We shall return to these results after seeing how genotype  $\times$  environment interactions can be detected and analyzed.

## Detection and Analysis

Provided genotypes ( $G$ ) are replicated in each environment as well as over environments ( $E$ ), genotype  $\times$  environment interactions ( $GE$ ) can be detected through a two-way analysis of variance, and their relative importance determined from the components of variance:

$$\sigma_P^2 = \sigma_G^2 + \sigma_E^2 + \sigma_{GE}^2 + \sigma_e^2$$

The observed variance of genotype means over  $n$  environments with  $r$  replicates in each environment is:

$$\sigma_G^2 + \sigma_{GE}^2/n + \sigma_e^2/(nr)$$

The standard error of the difference (SED) between two genotype means is:

$$\text{SED} = (\text{2VAR})^{\frac{1}{2}}$$

where

$$\text{VAR} = \sigma_{GE}^2/n + \sigma_e^2/(nr)$$

The critical difference ( $d_\alpha$ ) between two genotypes is the difference that can occur by chance in a proportion  $\alpha$  of trials when the true difference is zero.

$$d_\alpha = t_\alpha \times \text{SED} \quad (\text{or as a percentage of the mean yield of the trials } \mu; d_\alpha = t_\alpha \times \text{SED} \times 100/\mu)$$

where  $t_\alpha$  is a value from the normal distribution that is exceeded with probability  $\alpha$ .

Potential new cultivars are usually evaluated in a series of trials conducted in an appropriate range of locations ( $L$ ) for more than 1 year ( $Y$ ), before making recommendations to farmers and end users. Locations will differ in factors such as soil type and agronomic management practices, as well as latitude, altitude and climate, and years will differ primarily in weather. Differences in disease levels may occur both between locations and years.

The standard error of the difference between the means of a new genotype and  $c$  checks evaluated in trials with  $r$  replicates at  $l$  locations for  $y$  years is:

$$\text{SED} = [(1 + c)\text{VAR}/c]^{\frac{1}{2}}$$

where

$$\text{VAR} = \sigma_{GL}^2/l + \sigma_{GY}^2/y + \sigma_{GLY}^2/(ly) + \sigma_e^2/(lyr)$$

Estimates of the components of variance from a series of trials can be used to evaluate the efficiency and effectiveness of the trials system in terms of the number of locations, years and replicates used; and to make improvements where necessary (Robinson 1984; Talbot 1984, 1997). The residual (or restricted) maximum likelihood (REML) approach to estimation can be used on incomplete ‘genotypes  $\times$  locations  $\times$  years’ data matrices (Talbot 1997).

Data from a set of trials will help the breeder to choose cultivars for use as parents in further crosses; but at the start of a programme additional germplasm of interest should also be included. The breeder will also want to know if there are changes in the ranking of cultivars such that one cultivar does consistently better than another in one environment but consistently worse in another environment. Baker (1988) has shown how a technique developed by Azzalini and Cox (1984) can be used for this purpose.

### ***Yield Potential and Drought Resistance or Tolerance***

Crops can cope with drought by three mechanisms, namely escape, avoidance and tolerance. Thus upland rice can escape drought by early maturity, avoid it through drought-induced elongation of roots to reach comparatively deeper moisture zones, or tolerate it through reducing transpiration losses by leaf rolling, early closure of stomatal openings and cuticular resistance (Virmani and Ilyas-Ahmed 2007). Blum (2005) has given cereals grown under conditions of variable water supply as a good example of crossover G  $\times$  E (Fig. 7.1) with a physiological explanation. He points out that for most cereals grown under water-limited conditions the crossover occurs at a yield level of around 2–3 t/ha, which is approximately one-third of the yield potential. The main reason for a crossover under conditions of variable water supply is an inherent difference among the tested cultivars in drought resistance, beyond the difference in their yield potential. Blum (2005) then offers a possible explanation in more detail. Effective and successful selection for yield under water-stress most likely results in a dehydration-avoidance (rather than tolerance) phenotype, which is characterized by the maintenance of high plant water status under stress. Such a phenotype can be achieved by early flowering, a smaller plant, a smaller leaf area and limited tillering (in cereals), all of which are in contrast to a

high yield potential phenotype. Selection for high yield potential in cereals means a large sink demand (the developing grains), which constitutes a load on the shoot in terms of its water status and turgor maintenance under drought stress. This is the result of the strong assimilate export from the leaves into the grain and the increase in stomatal conductance. Therefore, a crossover interaction for yield is to be expected over a range of environments.

Blum (2005) also points out that crossover interaction can theoretically be avoided if high yield potential can be combined with drought resistance through selection of dehydration-avoidance factors that are not associated with lower yield potential. He puts forward two possibilities for consideration, namely insensitivity to the hormone abscisic acid (ABA) and osmotic adjustment, which has no obligatory association with low yield potential. He also warns against thinking you simply have to select for greater water use efficiency (WUE) by carbon-isotope discrimination under various levels of water supply. This is because WUE is largely a function of reduced water use rather than a net improvement in plant production or the biochemistry of CO<sub>2</sub> assimilation (WUE = Yield/Water Use). WUE on the basis of reduced water use is expressed in improved yield under water-limited conditions only when there is a need to balance crop water use against a limited and known soil moisture reserve. However, under most dryland situations where crops depend on unpredictable seasonal rainfall, the maximization of soil moisture use is a crucial component of drought resistance (avoidance), which is generally expressed in lower WUE. For example, a greater capacity for water absorption may be accomplished through longer, branched roots which can exploit deeper soil water reserves. The interested reader can find more information on drought resistance by dehydration avoidance or tolerance, and on WUE, in the review by Blum (2005) and the book edited by Fritsche-Neto and Borém (2012).

In conclusion, selection for wide adaptability will usually involve selection in more than one environment and for more than one physiological component (mechanism). For example, the CIMMYT strategy for spring wheat is to ensure that drought-tolerant germplasm maintains responsiveness if more moisture becomes available in a season (Singh and Trethowan 2007). Segregating populations are grown in alternate generations under favourable environments and drought-stressed environments to combine high yield potential and drought-stress tolerance. Furthermore, a drip-irrigation system is used in CIMMYT's experimental field in northwestern Mexico (usually negligible rainfall during crop season) so that exact amounts of water can be applied at chosen growth stages to generate different drought scenarios which represent different parts of the world. Lines can also be selected for adaptation to heat by late planting. Finally, it is worth pointing out that there is still much interest in the stay-green (delayed senescence) trait in sorghum as a means of providing post-flowering, end of growing season, drought tolerance. For example, in post-rainy crop seasons in India, Reddy et al. (2014) did a QTL analysis of 245 F<sub>9</sub> recombinant inbred lines from the cross between a widely adapted cultivar M35-1 and the popular stay-green genotype B35. They identified *Stg2*, *Stg3* and *StgB* as key QTL for marker-assisted selection to improve terminal drought tolerance.

## Salt Tolerance and Yield

Soil salinity is another major environmental constraint to crop production. As mentioned in Chap. 2, Qadir et al. (2014) estimated that by 2014 about 62 million hectares (20 %) of the world's irrigated lands had been degraded by salt, up from 45 million in the early 1990s. This area is expected to increase due to global climate change and as a consequence of many irrigation practices (Roy et al. 2014). The driving force for the movement of water in plants is a water-potential gradient generated by transpiration, as explained in textbooks of botany (Moore et al. 1998). Water flows passively from areas of high to low (water) potential energy. The presence of salts in soil decreases its water potential so that in extreme situations water moves from roots to soil instead of from soil to roots. Likewise as soil dries, it water potential decreases to the point where water moves from root hair to soil. One other key point is that the solute (osmotic) component of water potential is proportional to the number of dissolved particles and is independent of type of particle. The more particles that are present, the lower the water potential.

Roy et al. (2014) have reviewed the genetic basis of salt tolerance and possible combinations of genes that may need to be introduced into elite crop cultivars. Briefly they make the following points. Within minutes of application of salt in an experimental system one observes stomatal closure, with concomitant increases in leaf temperature and inhibition of shoot elongation. This is often referred to as the osmotic phase because the primary effect is on water potential, preventing water uptake by roots and water efflux from cells. The primary outcome is an overall reduction in production of new leaves and a significant reduction in shoot growth. However, a slower onset inhibition of growth (and reproductive development) occurs over several days to weeks and is due to accumulation of salt, especially in older leaves, causing the premature senescence of those older leaves. This is due to both the accumulation of salts, and the inability of the shoot to tolerate the salt that has accumulated to toxic concentrations; this is referred to as the ionic phase. In nature, plants that are sensitive to high-salt conditions (glycophytes) commonly respond to salt stress by accumulating osmoprotectants, whereas plants that can tolerate high-salt conditions (halophytes) employ specific mechanisms to avoid the toxic effects of  $\text{Na}^+$  (sodium) and  $\text{Cl}^-$  (chloride) ions by transporting them out of the cytoplasm. Since the transport works against a concentration gradient, it requires energy which is achieved by coupling the transport (antiport) protein to a proton ( $\text{H}^+$ ) pump working in the opposite direction. Details can be found in Slater et al. (2008), but let us return to Roy et al. (2014).

The salinity tolerance of crops can be improved by conventional breeding through the location of QTLs for components of tolerance and then marker-assisted selection. It can also be improved by genetic transformation through the incorporation of genes for osmotic tolerance, ion exclusion and tissue tolerance. Osmotic tolerance, which minimizes the reduction of shoot growth, is poorly understood but thought to be regulated by long distance signals that are triggered before shoot  $\text{Na}^+$  accumulation. In ion exclusion,  $\text{Na}^+$  and  $\text{Cl}^-$  transport processes in roots reduce the

accumulation of toxic concentrations of  $\text{Na}^+$  and  $\text{Cl}^-$  within leaves. Mechanisms may include retrieval of  $\text{Na}^+$  from the xylem, compartmentation of ions in vacuoles of cortical cells and/or efflux of ions back to the soil. In tissue tolerance, high salt concentrations are found in leaves but are compartmentalized at the cellular level, especially in the vacuole. Synthesis of compatible solutes and production of enzymes catalyzing detoxification of reactive oxygen species (cause oxidative stress) also occur. Compatible solutes such as glycine betaine, mannitol, proline and trehalose are non-toxic compounds that reduce osmotic potential and hence osmotic stress, thus acting as osmoprotectants. Although  $\text{Na}^+$  and  $\text{Cl}^-$  also reduce osmotic potential, they disrupt the water shell around proteins and therefore denature them. Osmoprotectants also overcome the osmotic stress generated by water stress.

Roy et al. (2014) present a compilation of transgenes and their effects on salt tolerance traits in crops and explain that temporal and spatial (specific cell types) control of gene-expression are often important for increasing tolerance. Importantly, the effects of introduced genes need to be evaluated in the field to determine their effect on salinity tolerance and yield improvement. Hence once again we have the issue in breeding strategy of how tolerance can be combined with high yield potential in the absence of stress. An example comes from the most intensively studied component of tolerance, namely the exclusion of  $\text{Na}^+$  from leaf blades; a trait that is relatively straightforward to phenotype. Munns et al. (2012) demonstrated that wheat grain yield on saline soils is improved by an ancestral  $\text{Na}^+$  transporter gene. They introgressed a source of  $\text{Na}^+$  exclusion (*Nax2*) from diploid *Triticum monococcum* (A genome) into tetraploid durum wheat cultivar Tamaroi, and then produced near-isogenic lines with and without *Nax2*. The *Nax2* locus confers a reduced rate of  $\text{Na}^+$  transport from roots to shoots by retrieving  $\text{Na}^+$  from the root xylem. The researchers provided compelling evidence that *TmHKT1;5-A* is the gene responsible for the  $\text{Na}^+$  shoot-exclusion phenotype controlled by the *Nax2* locus. *TmHKT1;5-A* encodes a high-affinity  $\text{Na}^+$ -specific transporter located on the plasma membrane. In field trials in Australia, lines containing *TmHKT1;5-A* yielded significantly more than Tamaroi in high salinity ( $0.3 \text{ t ha}^{-1}$  or 24 % more) through more grains per plant rather than larger grains. Lines containing *TmHKT1;5-A* had similar yields to the recurrent parent Tamaroi in less saline soil; there was no 'yield penalty' associated with the presence of *TmHKT1;5-A*.

## Regression Analysis

Sometimes all or most of the genotype  $\times$  environment interaction variance can be accounted for by the heterogeneity of regression ( $\beta_{ij}$ ) in a joint regression analysis of the interactions on to an environmental index. In the absence of a physical assessment of the environments, the mean value of all genotypes in a given environment ( $E_j$ ) is used as a biological assessment of that environment (Finlay and Wilkinson 1963), and is expressed as a deviation from the overall mean ( $\mu$ ) where  $\bar{e}_{ij}$  is the average plot error for  $r$  replicates.

$$\begin{aligned} P_{ij} &= \mu + G_i + E_j + (GE)_{ij} + \bar{e}_{ij} \\ &= \mu + G_i + E_j + \beta_i E_j + d_{ij} + \bar{e}_{ij} \end{aligned}$$

In our simple example (Fig. 7.1 and Table 7.1) the heterogeneity of regressions accounts for all of the interaction (i.e. the  $d$ 's all equal zero) and we can see that genotypes 2 and 3 have regressions with slope 1 ( $\beta = 0$ ), whereas genotype 1 has slope 0 ( $\beta = -1$ ) and genotype 4 has slope 2 ( $\beta = 1$ ). Under these circumstances the regression coefficients are properties of the genotypes *per se* and measure their responsiveness (sensitivity) to environmental change, genotype 4 being the most responsive and genotype 1 the least responsive.

**Table 7.1** Analysis of yields of four genotypes grown in three environments

(a) Yields ( $Y_{ij}$ ) of four genotypes grown in three environments where  $Y_{ij}$  is yield of  $i$ th genotype in  $j$ th environment

	Environment 1	Environment 2	Environment 3	Mean
Genotype 1	8	8	8	8
Genotype 2	7	9	11	9
Genotype 3	5	7	9	7
Genotype 4	4	8	12	8
Mean	6	8	10	8

(b)  $G \times E$  ( $Z_{ij}$ ) for four genotypes in three environments ( $Z_{ij} = Y_{ij} - Y_i - Y_j + Y_{..}$ ), where  $Y_i$  is the mean yield of genotype  $i$ ,  $Y_j$  is the mean yield of environment  $j$  and  $Y_{..}$  is the overall mean yield; together with  $\beta_i$ , where  $(1 + \beta_i)$  is the slope of the regression line for genotype  $i$  in Fig. 7.1, and the principal component scores for genotypes and environments

	Environment 1	Environment 2	Environment 3	$\beta_i$	PCA 1 scores
Genotype 1	2	0	-2	-1	$-2/\sqrt{2}$
Genotype 2	0	0	0	0	0
Genotype 3	0	0	0	0	0
Genotype 4	-2	0	2	1	$2/\sqrt{2}$
$E_j$	-2	0	2		
PCA 1 scores	$-2/\sqrt{2}$	0	$2/\sqrt{2}$		

(c) Analysis of variance

Source	df	SS	MS
Environment	2	32	16
Genotype	3	6	2
$G \times E$	6	16	$2\frac{2}{3}$
Regression	3	16	$5\frac{1}{3}$
Residual	3	0	0
$G \times E$			
PCA 1	4	16	4
PCA 2	2	0	0
Total	11	54	

## Selecting Genotypes in Different Types of Environment

A consideration of the sensitivity of genotypes to environmental change has led to predictions about the consequences of selecting genotypes in different types of environment. Upward selection in above-average environments and downward selection in below-average environments should result in genotypes with a high sensitivity, whereas upward selection in below-average environments and downward selection in above-average environments should result in ones with a low sensitivity. These predictions have been confirmed in selection experiments with the fungus *Schizophyllum commune* (Jinks and Connolly 1973, 1975) and with the tobacco *Nicotiana rustica* (Jinks and Pooni 1982). Furthermore, Messmer et al. (2012) found substantial evidence in the literature that breeding cereals for low N input conditions is more efficient under severe N stress than under high-input conditions. Nevertheless, there are examples where high yielding cultivars have been bred which do not display genotype  $\times$  environment interactions when compared with older, lower yielding ones. Bingham (1979) reported an experiment in which the yields of his highest yielding winter wheat cultivars were 40–50 % above the older ones, Little Joss and Holdfast. With one exception, the new cultivars did not benefit more than the old ones from greater nitrogen supply (105 versus 40 kg N/ha). The higher yield of the new cultivars was strongly correlated with an increase in harvest index, the ratio of grain yield to grain plus straw yield at maturity. In potatoes, Simmonds (1981) analyzed the increase in agricultural yields in Great Britain from 34 t/ha in 1964 to 42.9 t/ha in 1976. He concluded that 4.0 t/ha was environmental (better husbandry and management), 5.5 t/ha was due to the near replacement of cultivars ‘King Edward’ and ‘Majestic’ with three more modern cultivars ‘Pentland Crown’, ‘Maris Piper’ and ‘Desiree’, and that the difference of 0.6 t/ha was a small but negative genotype  $\times$  environment effect. In other words, the new cultivars were not more responsive than the old ones to environmental change. Simmonds (1981) found similar results for historically rising yields in sugarcanes; but in contrast for various cereals, both temperate and tropical, evidence that past selection had produced responsive varieties with genotype  $\times$  environment interaction effects contributing about one third of the estimated yield increase. Clearly the extent and importance of genotype  $\times$  environment interactions in any given crop situation need to be determined by experiment.

## Response Curves and Deviations from Regression

When the heterogeneity of regression accounts for most but not all of the interaction, the regression coefficients still have a predictive value, but they are no longer properties of the genotypes *per se*, nor are the deviations from regression. These points can be seen by considering the response curves and surfaces of genotypes to changes in their physical environment, and translating them into regression lines (Knight 1970, 1973). A simple response curve is as follows (Knight 1973):

$$y = a + bx - cx^2,$$

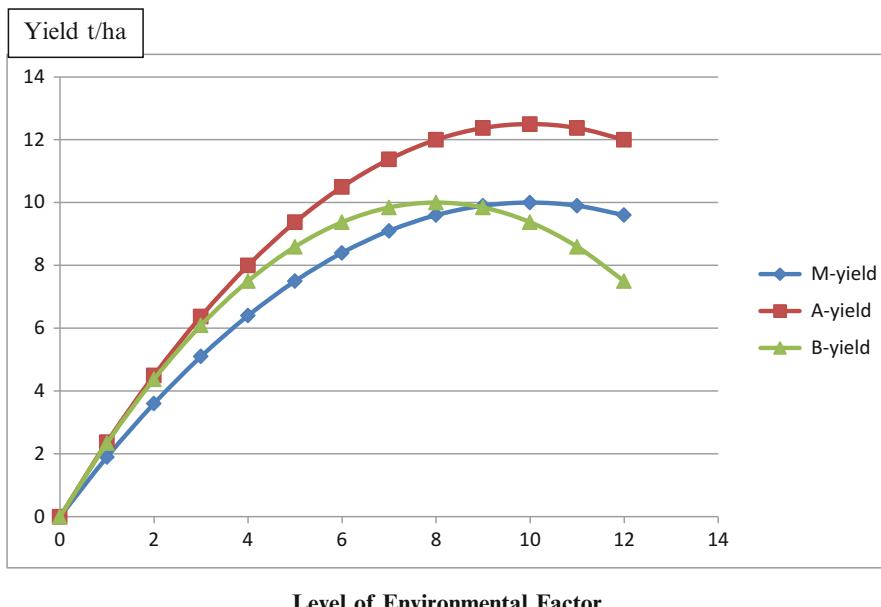
where  $y$  is the yield response,  $x$  the level of one environmental factor and the coefficients  $a$ ,  $b$  and  $c$  are constants for any one genotype. Hence

$$y_{ij} = a_i + b_i x_j - c_i x_j^2$$

where  $y_{ij}$  is the yield of genotype  $i$  in environment  $j$  and  $x$ .

Response curves have been worked out for the application of fertilizers to crops so that recommendations can be made to farmers about the amounts to apply for maximum yield and the yield reductions likely from saving money by applying lower amounts. Let us consider possible curves for nitrogen fertilizer.

If  $M$  is the response curve for the mean of many genotypes (Fig. 7.2), then in a Finlay and Wilkinson (1963) regression analysis, it has a line of slope 1, with no deviations, and passes through the origin. We now consider the regression analysis for genotypes A and B, where  $M$  is the biological environmental index (i.e. the  $x$ -values). We ignore the curves above environmental level 10 on the grounds that all genotypes display a reduction in yield and hence this is the maximum level of interest. Genotype A has a higher maximum yield than  $M$  but at the same level of the physical index. Its regression line has a slope of 1.25 ( $\delta_A = 0.25$ ) with no deviations. In contrast, genotype B has the same maximum yield as  $M$  but at a lower level of the physical index. Its regression line has a slope of 0.98



**Fig. 7.2** Yield responses of genotypes A, B and M (mean of all genotypes) to an environmental factor such as nitrogen fertilizer

( $b_A = -0.02$ ), which is very close to 1, but there are also deviations from regression. The regression line accounts for 97 % of the total sum of squares for genotype B over environments, and hence has good predictive ability. But it would not make sense to say that B is less stable than A; it is simply different (Lin et al. 1986). In fact the total sum of squares for A (185.3) is higher than that for B (117.6), reflecting its greater variability. The regression analysis would reveal that genotype A has the highest yield at all fertilizer levels, but fail to reveal that genotype B requires less fertilizer for maximum yield than the average over all genotypes. Now is an opportune moment to briefly consider nitrogen use efficiency; and then as a contrast, phosphorus use efficiency.

### **Nitrogen Use Efficiency**

Today in many crops, high yields are achieved through the application of high levels of nitrogen fertilizer. However, its manufacture using the Haber-Bosch process is expensive and further costs come from transporting and applying the synthetic fertilizer. Furthermore, globally, only a third of the nitrogen applied to cereal crops as fertilizer is harvested in the grain (Raun and Johnson 1999). Excess fertilizer has adverse effects on the environment in terms of water, air and soil quality. Understandably then, breeding cultivars with increased nitrogen use efficiency (NUE) makes sense, but is complicated, as explained by Dawson et al. (2008) and Xu et al. (2012) in their reviews. Nitrogen use involves nitrogen uptake from the soil followed by nitrogen translocation, assimilation and remobilization (during leaf senescence) within the plant. Each step is governed by many interacting genetic and environmental factors (Xu et al. 2012). Overall NUE is not simply the yield achieved divided by the amount of applied nitrogen fertilizer. Not all of the applied nitrogen is available to the plant nor is it the only source of available nitrogen. Furthermore, with low-input and organic farming systems the inputs of synthetic nitrogen are minimized or absent. Dawson et al. (2008) explain how NUE can be defined to distinguish between soil and plant physiological processes and effects. Briefly, for cereal grain, NUE ( $G_w/N_s$ ) is grain dry weight ( $G_w$ ) divided by the supply of nitrogen ( $N_s$ ). Plant available nitrogen ( $N_{av}$ ) is the difference between nitrogen supply and losses from leaching, volatilization, runoff, denitrification and immobilization. Hence plant available NUE is  $G_w/N_{av}$ , which is a measure of efficiency for the plant, whereas  $G_w/N_s$  is the efficiency of the whole plant/soil system. Two plant physiological components contribute to overall NUE, nitrogen uptake (acquisition) efficiency ( $N_t/N_s$ , where  $N_t$  is total nitrogen in plant) and nitrogen utilization efficiency ( $G_w/N_t$ ). As before, available nitrogen uptake efficiency ( $N_t/N_{av}$ ) and utilization efficiency contribute to available NUE. Uptake efficiency is therefore a measure of how much nitrogen the plant absorbs in proportion to either the nitrogen supply or the plant available nitrogen, and is associated with the assimilation of nitrogen compounds into plant tissues. Utilization efficiency measures the response of grain yield to the total

nitrogen in the plant. More generally, resource use efficiency always has two components: acquisition efficiency (ability to absorb resource) and utilization efficiency (ability to produce biomass or grain with absorbed resources).

The aim of the bread wheat programme at CIMMYT is to make improvements in both uptake efficiency (UPE) and utilization efficiency (UTE) in order to maintain the wide adaptation of its germplasm (Van Ginkel et al. 2001). CIMMYT have produced cultivars which outyield earlier semi-dwarf wheats as well as the old tall cultivars under both low and high nitrogen levels. Selection was practised under intermediate levels of nitrogen fertility (100–200 units/ha N) as both UPE and UTE are expressed and can be selected. However, Van Ginkel et al. (2001) have shown in Mexico that alternating between high (300 kg/ha N) and low (none added) N from  $F_2$  to  $F_6$ , but not the reverse, resulted in higher yields at intermediate and high N levels than was achieved by always selecting under low, medium or high N. The selection practised was visual selection for superior agronomic type. Increased yields from alternating between high and low N were due to increased biomass (not harvest index) and were accompanied by higher UPE values; that is, increased ability to move nitrogen from the soil into the plant. All selection regimes produced similar improvements in yields at low N levels. Van Ginkel et al. (2001) generated the segregating material by crossing two lines with high N uptake efficiency (UPE) with two lines with high N utilization efficiency (UTE).

Where protein content is as important as yield, grain nitrogen accumulation efficiency ( $GN_{ace} = N_g/N_s$ , where  $N_g$  is the nitrogen in the grain) and nitrogen harvest index ( $N_g/N_t$ ) will also be considerations. The former is a measure of the overall efficiency with which plants extract nitrogen from the soil and accumulate it in the grain by harvest, and the latter is a measure of nitrogen translocation efficiency. In fact, Xu et al. (2012) consider increasing both the grain and nitrogen harvest index, to drive nitrogen acquisition and utilization, are important approaches for breeding future high-NUE cultivars.

These definitions help to clarify the components of NUE and hence possible physiological selection criteria to use in breeding programmes. However, as with water use, breeders still need to consider whether they can breed cultivars that are superior to others under both conventional farming systems, with high inputs of synthetic nitrogen fertilizer, and low-input and organic farming systems with their different nitrogen sources, cycling and management strategies. Dawson et al. (2008) suggested that it may be possible to combine high nitrogen productivity for the former with higher levels of internal nitrogen conservation for the latter. They thought that beneficial genetic traits might include the ability to maintain photosynthesis and nitrogen uptake under nitrogen stress and the ability to extract soil nitrogen at low concentrations, perhaps through beneficial associations with soil micro-organisms. Xu et al. (2012) also concluded that the limiting factors in plant metabolism for maximizing NUE are different at high and low nitrogen supplies, indicating great potential for improving the NUE of the many current cultivars that were bred in well-fertilized soil. They also pointed out that high yields require the coordination of carbohydrate and nitrogen metabolism.

## ***Phosphorus Use Efficiency***

The use of phosphate fertilizer to increase crop yields is complicated by the ability of soils to adsorb applied phosphorus, making it unavailable to the crop. However, during the course of the growing season, the mineralization (hydrolyzation) of any organic matter present results in a slow release of inorganic phosphate into the soil solution, thus making this source of phosphorus available for absorption by plant roots. Furthermore, available phosphorus content becomes higher and higher in the most cultivated soils with intensive cropping systems, particularly in the soil-surface zone, as a result of long-term application of phosphorus fertilizer and long-term organic residue accumulation (Wang et al. 2010). Nevertheless, plants must absorb phosphate from dilute solutions and the size and area of their root system are key points in determining their ability to acquire phosphate from the soil by diffusion. Therefore breeding crop cultivars that can grow and yield better with low phosphorus supply is important for improving crop production. Improving phosphate use efficiency can be achieved both through increases in acquisition efficiency and increases in utilization efficiency; but their relative contributions need to be determined experimentally for any particular crop species, set of cultivars and environmental conditions, such as soil phosphorus status. In their review, Wang et al. (2010) give crop examples (cereals and beans) where phosphorus supply was limited and acquisition efficiency contributed most to phosphorus use efficiency, and other examples (wheat and potatoes) where phosphorus supply was adequate and it was utilization efficiency that contributed most to phosphorus use efficiency. They also present evidence in soybean that modern cultivars have been selected to acquire more phosphorus from the phosphorus-rich soil surface zone. The modern cultivars have shallow root architecture and high acquisition efficiency whereas wild genotypes have deep root architecture and low acquisition efficiency.

Wang et al. (2010) go on to argue that acquisition efficiency has already been selected in modern cultivars. Therefore utilization efficiency is likely to be the significant bottleneck for further improvements in phosphorus use efficiency in modern crops grown in intensive cropping systems, such as hybrid rice. They also point out that the modification of root systems usually requires additional carbon input so that crops might have to sacrifice carbohydrates that could be used for higher yields to meet the demand for phosphorus acquisition. However, acquisition efficiency may be the most important consideration when phosphorus is in limited supply. For example, Gamuyao et al. (2012) have shown that constitutive overexpression of the gene *PSTOL1* in modern rice varieties (IR64 and Nipponbare), which are intolerant of phosphorus starvation, significantly enhances grain yield in phosphorus-deficient soil. Further analyses showed that *PSTOL1* acts as an enhancer of early root growth, thereby enabling plants to acquire more phosphorus and other nutrients. The gene was initially discovered as a major QTL (*Pup1*) for phosphorus-deficiency tolerance in the traditional rice cultivar Kasalath, which originates from a region in India with poor soils. Sequencing the *Pup1* locus in Kasalath revealed the presence of a 90 kb transposon-rich insertion-deletion (indel) that was absent from the Nipponbare reference genome. The most

likely candidate gene was a *Pup1*-specific protein kinase gene. The finding that root growth was enhanced in *PSTOL1*-overexpressing lines as well as in *Pup1* introgression lines provided strong evidence that *PSTOL1* was the major tolerance gene within the *Pup1* QTL and that this gene acts at least partially independently of phosphorus supply. Introduction of the gene, by introgression or genetic transformation, into locally adapted rice varieties in Asia and Africa is expected to considerably enhance productivity under low phosphorus conditions.

At the start of a breeding programme, when determining possibilities, it would seem worthwhile assessing modern cultivars, landraces and wild relatives of crop plants for yield and phosphorus use efficiency under limited and adequate supply of phosphorus. This way the breeder can find out if genotypes exist, or can be bred, that combine acquisition and utilization efficiency. Such genotypes should yield well under different phosphorus supply conditions in different soil types. The breeder can then seek to discover the relevant genes. The same is true when breeding for any resource use efficiency and any tolerances to stresses. Examples of breeding for nitrogen and phosphorus use efficiency, together with breeding for water use efficiency and tolerance to salinity, aluminium toxicity and heat-stress, can be found in *Plant Breeding for Abiotic Stress Tolerance*, edited by Fritsche-Neto and Borém (2012). The book also covers the physiological mechanisms involved.

## Principal Component Analysis

The contribution of the heterogeneity of regressions to the interaction is in the form of a genotypic contribution ( $\theta_i$ ) multiplied by an environmental contribution ( $E_j$ ) (Table 7.1). In general, the complexity or otherwise of genotype  $\times$  environment interactions can be seen from a principal component analysis (PCA). If the overall (additive) effects of genotypes and environments are first removed from the data (Table 7.1), the analysis results in multiplicative terms for the interaction, the so called additive main effects and multiplicative interaction (AMMI) model (Gauch 1988; Gauch et al. 2008):

$$P_{ij} = \mu + G_i + E_j + \sum \theta_k u_k v_{kj} + \bar{e}_{ij},$$

where summation is from  $k = 1$  to  $K$ , the number of PCA axes retained in the model, and  $\theta$ ,  $u$  and  $v$  are the singular values (i.e. square roots of the eigenvalues), normalized PCA scores for genotypes and normalized PCA scores for environments, respectively. Axis  $k$  is assigned  $(G+E-1-2k)$  degrees of freedom, where  $G$  and  $E$  are the number of genotypes and environments, respectively. When considering yield data,  $P_{ij}$  is often replaced  $Y_{ij}$ , the yield of  $i$ th genotype in  $j$ th environment. The first and second principal components may account for most of the interaction and may have a biological meaning, in which case the nature of the interaction can be visualized in biplots (Kempton 1984). If the first principal component (PC) explains all or most of the variation, the PC1 scores for genotypes and environments are plotted (y-axis) against the additive means for genotypes ( $Y_i$ ) and environments ( $Y_j$ ) and the overall mean ( $Y..$ ) (x-axis); whereas if two principal

components are required, the PC2 scores are plotted against the PC1 scores. The scores are usually scaled by multiplying by  $\theta^{1/2}$  to give  $\theta^{1/2}u$  and  $\theta^{1/2}v$  so that  $\Sigma \theta_{ki} u_i v_{kj}$  becomes  $\Sigma \theta_{ki}^{1/2} u_i \theta_{kj}^{1/2} v_{kj}$ . The breeder can thus identify interesting interactions and hence genotypes and possible aspects of the environment (e.g. drier conditions or poorer soil) for further research.

In our simple example in Table 7.1, all the genotype  $\times$  environment interactions can be explained by the first principal component. This is most easily seen using matrix algebra. If our  $G \times E$  matrix  $Z$ , with elements  $z_{ij}$ , is post multiplied by its transpose, we obtain the matrix required to determine the PCA scores for genotypes, as well as the eigenvalues:

$$\mathbf{ZZ}' = \begin{matrix} 8 & 0 & 0 & -8 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -8 & 0 & 0 & 8 \end{matrix}$$

The trace of the matrix (sum of the elements in the leading diagonal) is 16 which is the sum of squares for  $G \times E$  ( $z_{11}^2 + z_{12}^2 + \dots + z_{43}^2$ ).

We now consider the following set of homogeneous equations:

$$(\mathbf{ZZ}' - \lambda \mathbf{I})\mathbf{U} = \mathbf{0},$$

where  $\mathbf{I}$  is the unit matrix of the same order as  $\mathbf{ZZ}'$  and the  $\lambda$ s are the eigenvalues of  $\mathbf{ZZ}'$  for which the corresponding  $\mathbf{U}$ s, called the eigenvectors, are the non-trivial solutions (i.e.  $\mathbf{U} \neq \mathbf{0}$ ) of the equations which occur when the determinant  $|\mathbf{ZZ}' - \lambda \mathbf{I}| = 0$ .

For our set of equations there is just one non-trivial solution which occurs when  $\lambda = 16$ .

With just one eigenvalue, its value (16) is the trace of matrices  $\mathbf{Z}'\mathbf{Z}$  and  $\mathbf{ZZ}'$ , the sum of squares for  $G \times E$ . More generally, the sum of the eigenvalues of a matrix is equal to its trace.

The associated normalized (squares of elements sum to 1) eigenvector (scores) for genotypes is:

$$\begin{matrix} -1/\sqrt{2} \\ 0 \\ 0 \\ 1/\sqrt{2} \end{matrix}$$

This is usually scaled by multiplying by  $\theta^{1/2}$  ( $= \lambda^{1/4} = 2$ ) (as in Table 7.1(b)):

$$\begin{matrix} -2/\sqrt{2} \\ 0 \\ 0 \\ 2/\sqrt{2} \end{matrix}$$

If our  $G \times E$  matrix  $\mathbf{Z}$  is pre-multiplied by its transpose, we obtain the matrix required to determine the PCA scores for environments, as well as the same eigenvalues.

$$\mathbf{Z}'\mathbf{Z} = \begin{matrix} 8 & 0 & -8 \\ 0 & 0 & 0 \\ -8 & 0 & 8 \end{matrix}$$

As just mentioned, the trace of this matrix is again 16, the sum of squares for  $G \times E$ .

There is just one non-trivial solution of the equation  $(\mathbf{Z}'\mathbf{Z} - \lambda\mathbf{I})\mathbf{V} = \mathbf{0}$ , which occurs when  $\lambda = 16$ , and the associated normalized eigenvector (scores) for environments is:

$$\begin{matrix} -1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{matrix}$$

This is usually scaled by multiplying by  $\theta^{1/2}$  ( $=\lambda^{1/4}=2$ ) (as in Table 7.1(b)):

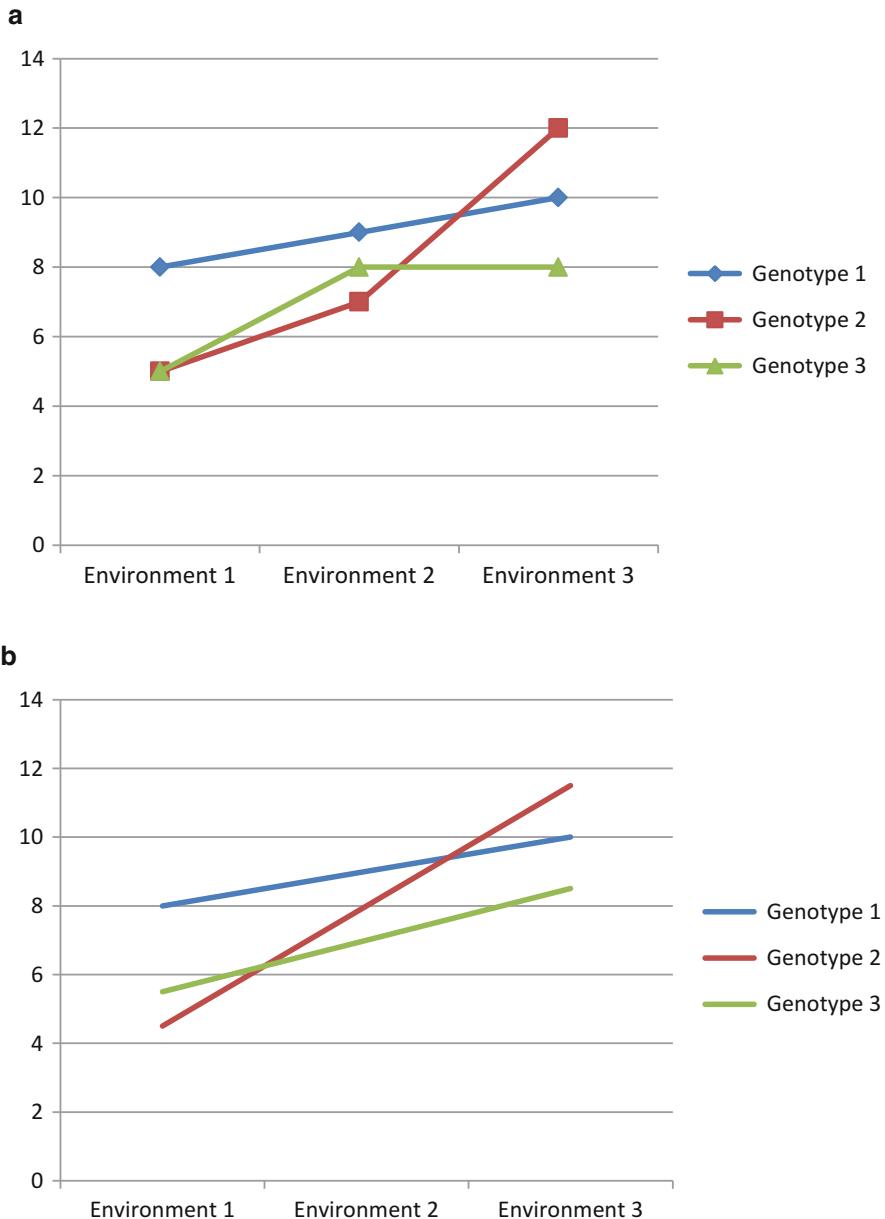
$$\begin{matrix} -2/\sqrt{2} \\ 0 \\ 2/\sqrt{2} \end{matrix}$$

The rescaled genotype scores multiplied by the transpose of the rescaled environment scores reproduce the genotype  $\times$  environment interactions:

$$\begin{pmatrix} -2/\sqrt{2} & -2/\sqrt{2} & 2/\sqrt{2} \end{pmatrix} = \begin{pmatrix} 2 & 0 & -2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} -2 & 0 & 2 \end{pmatrix}$$

They could however have been rescaled in such a way that they have the same values as  $\delta_i$  and  $E_j$ , thus confirming that the regression and PCA analyses give the same result when there is only one PCA axis.

A slightly more complicated example is given in Fig. 7.3, Table 7.2 and Box 7.1 where there are deviations from regression and two principal components are required in the PCA analysis. The first principal component accounts for 90 % of the  $G \times E$  sum of squares and provides a good prediction of the interactions, being out by only  $\frac{1}{2}$  in just four out of nine interactions. In contrast, the regression analysis accounts for 70 % of the sum of squares and the predictions are not quite so good, being out by 1 in two interactions in environment 2 and also getting the rankings of genotypes 2 and 3 wrong in this environment.



**Fig. 7.3** Analysis of yields of three genotypes grown in three environments. (a) Actual yields. (b) Regression analysis

**Table 7.2** Analysis of yields of three genotypes grown in three environments

(a) Yields ( $Y_{ij}$ ) of three genotypes grown in three environments where  $Y_{ij}$  is yield of  $i$ th genotype in  $j$ th environment

	Environment 1	Environment 2	Environment 3	Mean
Genotype 1	8	9	10	9
Genotype 2	5	7	12	8
Genotype 3	5	8	8	7
Mean	6	8	10	8

(b)  $G \times E$  ( $Z_{ij}$ ) for three genotypes in three environments ( $Z_{ij} = Y_{ij} - Y_{..} - Y_{i.} + Y_{..}$ ), where  $Y_{..}$  is the mean yield of genotype  $i$ ,  $Y_{j.}$  is the mean yield of environment  $j$  and  $Y_{..}$  is the overall mean yield; together with  $\delta_i$ , where  $(1 + \delta_i)$  is the slope of the regression line for genotype  $i$ ,  $\sum d_i^2$  the sum of squared deviations from regression for genotype  $i$  and the two principal component scores for genotypes and environments

	Environment 1	Environment 2	Environment 3	$\delta_i$	$\sum d_i^2$	PCA 1	PCA 2
Genotype 1	1	0	-1	-0.50	0.0	$-\sqrt{3}/\sqrt{6}$	$-1/\sqrt{2}$
Genotype 2	-1	-1	2	0.75	1.5	$2\sqrt{3}/\sqrt{6}$	0
Genotype 3	0	1	-1	-0.25	1.5	$-\sqrt{3}/\sqrt{6}$	$1/\sqrt{2}$
$E_j$	-2	0	2				
PCA 1	$-\sqrt{3}/\sqrt{6}$	$-\sqrt{3}/\sqrt{6}$	$2\sqrt{3}/\sqrt{6}$				
PCA 2	$-1/\sqrt{2}$	$1/\sqrt{2}$	0				

(c) Analysis of variance

Source	df	SS	MS
Environment	2	24	12
Genotype	2	6	3
$G \times E$	4	10	$2\frac{1}{2}$
Regression	2	7	$3\frac{1}{2}$
Residual	2	3	$1\frac{1}{2}$
$G \times E$			
PCA 1	3	9	3
PCA 2	1	1	1
Total	8	40	

### Box 7.1: Principal Component Analysis of Three Genotypes in Three Environments Shown in Table 7.2

$$\mathbf{Z}\mathbf{Z}' = \begin{pmatrix} 2 & -3 & 1 \\ -3 & 6 & -3 \\ 1 & -3 & 2 \end{pmatrix}$$

The trace of the matrix is 10 which is the sum of squares for  $G \times E$ .

There are two non-trivial solutions of the equation  $(\mathbf{Z}\mathbf{Z}' - \lambda\mathbf{I})\mathbf{U} = \mathbf{0}$ , with  $\lambda = 9$  and 1 (which sum to the trace = 10), and corresponding eigenvectors (scores) for genotypes:

(continued)

**Box 7.1** (continued)

$$\begin{array}{cc} \text{PCA1} & \text{PCA2} \\ (-1/\sqrt{6}) & (-1/\sqrt{2}) \\ (2/\sqrt{6}) & (0) \\ (-1/\sqrt{6}) & (1/\sqrt{2}) \\ \mathbf{Z}'\mathbf{Z} = & \begin{pmatrix} 2 & 1 & -3 \\ 1 & 2 & -3 \\ -3 & -3 & 6 \end{pmatrix} \end{array}$$

The trace of the matrix is 10 which is the sum of squares for  $G \times E$ .

There are two non-trivial solutions of the equation  $(\mathbf{Z}'\mathbf{Z} - \lambda\mathbf{I})\mathbf{V} = \mathbf{0}$ , with  $\lambda = 9$  and 1 (which sum to the trace = 10), and corresponding eigenvectors (scores) for environments:

$$\begin{array}{cc} \text{PCA1} & \text{PCA2} \\ (-1/\sqrt{6}) & (-1/\sqrt{2}) \\ (-1/\sqrt{6}) & (1/\sqrt{2}) \\ (2/\sqrt{6}) & (0) \end{array}$$

The contribution of PCA1 ( $\theta^{1/2} = \sqrt{3}$ ) to the genotype  $\times$  environment interactions is

$$\begin{array}{ll} (-\sqrt{3}/\sqrt{6})(-\sqrt{3}/\sqrt{6} & -\sqrt{3}/\sqrt{6} & 2\sqrt{3}/\sqrt{6}) = \\ (2\sqrt{3}/\sqrt{6}) & & (-1 & -1 & 2) \\ (-\sqrt{3}/\sqrt{6}) & & (\frac{1}{2} & \frac{1}{2} & -1) \end{array}$$

The contribution of PCA2 ( $\theta^{1/2} = 1$ ) to the genotype  $\times$  environment interactions is

$$\begin{array}{ll} (-1/\sqrt{2})(-1/\sqrt{2} & 1/\sqrt{2} & 0) = \\ (0) & & (0 & 0 & 0) \\ (1/\sqrt{2}) & & (-\frac{1}{2} & \frac{1}{2} & 0) \end{array}$$

The sum of the contributions from PCA1 and PCA2 equals the  $G \times E$  shown in Table 7.2.

In a real example, Zobel et al. (1988) reported the yield results for seven soybean cultivars grown in 35 site-year combinations in North America. Only the first principal component with 39 degrees of freedom was significant. In contrast, in Gauch's New York soybean trial with 15 genotypes and 15 environments, the first three principal components, with 27, 25 and 23 degrees of freedom, were significant

and accounted for 67 % of the sum of squares for G × E (Gauch 1988). As more components are added, the predicted G × E becomes closer to the observed G × E, but equally it is no longer possible to give simple biological interpretations of the interactions. However, the key messages for quantitative genetics and plant breeding can be seen in our simple example where all of the interaction is accounted for by heterogeneity of regressions. The plant breeder will often be seeking a cultivar that does best (or near to best) in average yielding environments and is close to the best cultivars in lower and higher yielding environments.

An alternative analysis of a set of yield trials combines and analyzes the total variation between genotype main effects and genotype × environment interactions (GGE model), but will not be considered here. The interested reader can consult the reviews in *Crop Science* mentioned earlier (Yan et al. 2007; Gauch et al. 2008).

## G × E as a Correlated Response to Selection

The concept of genetic correlation can also be applied to selection problems involving genotype × environment interactions by regarding a trait measured in two environments as two different, but correlated, traits. The improvement of performance in one environment, as a result of selection in another, is seen as a correlated response which depends on the square roots of the heritabilities of performance in each environment and the genetic correlation between the two performances (Falconer and Mackay 1996). We will consider responses to selection in detail in Chap. 10 and hence for now just state a theoretical result.

The correlated response in environment y ( $CR_y$ ) is the response in environment  $x$  ( $i_x h_x \sigma_{Gx}$ ) multiplied by the regression of genotypic values in  $y$  on those in  $x$ , the latter being the genetic correlation ( $r_G$ ) multiplied by  $\sigma_{Gy}/\sigma_{Gx}$ . In the current context we do not need to worry over whether the heritabilities are narrow or broad sense, and whether or not the genotypic effects are in fact breeding values (Chap. 10).

$$CR_y = i_x h_x \sigma_{Gx} \times r_G \sigma_{Gy} / \sigma_{Gx} = i_x h_x r_G \sigma_{Gy} = i_x h_x h_y r_G \sigma_{Py}$$

$i$  is intensity of selection

$h$  is square root of heritability

$r_G$  is genetic correlation

$\sigma_{Gy}$  is square root of genotypic variance in  $y$

$\sigma_{Py}$  is square root of phenotypic variance in  $y$ .

The response to direct selection in environment  $y$  ( $R_y$ ) is

$$R_y = i_y h_y \sigma_{Gy}$$

Hence the correlated response can be compared with the response ( $R_y$ ) expected from direct selection in environment  $y$ :

$$CR_y/R_y = r_G i_x h_x / i_y h_y$$

If the intensities of selection are similar, indirect selection in environment  $x$  will be superior to direct selection in environment  $y$  if the genetic correlation multiplied by the square root of the heritability in environment  $x$  is greater than the square root of the heritability in environment  $y$ .

### ***Conventional Versus Organic Farming***

A good example is breeding cultivars suitable for organic farming. Is organic farming so different from conventional farming that cultivars need to be bred especially for organic farming; for example, in order to cope with competition from weeds and limited nutrient availability from less readily available nutrient sources? Experimental data are needed to determine whether indirect (correlated response) selection under conventional farming ( $x$ ) for adaptation to organic farming is as effective as direct selection under organic farming ( $y$ ). The indirect selection could involve traits thought to be important for organic farming such as those capable of shading and hence suppressing weeds. If the same selection intensity is applied, the efficiency of indirect selection for a given trait depends on the heritabilities under the different farming systems and the genotypic correlation between the two farming systems. If the heritabilities are comparable under organic and conventional selection environments, then direct selection in the target environment is more efficient by an amount depending on how much  $r_G$  is less than one.

Only a few studies have been done to demonstrate the difference between direct and indirect selection for organic farming (Messmer et al. 2012), partly because a lot of work is involved. For example, Reid et al. (2011) assessed 79  $F_6$ -derived recombinant inbred lines of spring wheat in replicated yield trials on six organic and six conventional sites in diverse locations in Canada. The organic trial and conventional trial in 2005 were considered the selection trials, and the five organic and five conventional trials spread over 2006 and 2007 were considered the assessment trials. On average, grain yields were 53 % less in the organic trials and weed pressure was much greater. The tentative conclusion was that selection should be practised under organic farming because the top two lines for grain yield in the organic assessment trials ranked 31 and 38 in the conventional ones, and likewise the top two in the conventional assessments ranked 19 and 48 in the organic assessments. However, three out of the top eight in the organic assessment were also three out of the top eight in the conventional assessment. It is also of interest to note that selecting the top 8 out of the 79 lines in the conventional trial in 2005 resulted in a selection differential of 1.16 t/ha and a realized gain averaged over the subsequent five conventional trials of 0.46 t/ha (realized heritability of 0.40)

compared with 0.86 and 0.14 t/ha (realized heritability of 0.16) when the top eight were selected and assessed in the organic trials. In other words, breeding would appear more difficult for organic farming because of greater line by trial interactions. Hence organic breeding might be expected to benefit from participatory plant breeding with farmers selecting cultivars adapted to their own areas, regions, or even farms, rather than seeking more general adaptability; a consequence of which will be the diversity of cultivars desired by advocates of organic farming. Goldstein et al. (2012) reported the work mentioned by Messmer et al. (2012) on genetically broad populations of maize, strongly supporting direct selection under organic farming for complex traits like grain yield. In contrast, indirect selection was very efficient for highly heritable traits like dry matter content. The hybrid populations comprised 90 dihaploid flint lines crossed with a dent tester and 90 dihaploid dent lines crossed with a flint tester. For yield, heritability was almost the same under conventional ( $h^2 = 0.69$ ) and organic ( $h^2 = 0.61$ ) farming because for the latter, the genotypic variance and experimental error rate were both higher. However,  $r_G$  was 0.54 so that only about 50 % of the selection gain achievable under conventional conditions was also found under organic conditions.

### **Sorghum in West Africa**

Another good example comes from sorghum (*Sorghum bicolor*) breeding in West Africa, where the crop is generally cultivated with limited or no fertilization on soils of low phosphorus availability. Leiser et al. (2012) assessed grain yields of 70 diverse sorghum genotypes under –P (no P fertilization) and +P conditions at two locations in Mali over 5 years. Delayed heading dates (0–9.8 days) and reductions of grain yield (2–59 %) and plant height (13–107 cm) were observed in –P relative to the +P trials. Estimates of genetic variance and broad-sense heritabilities were high for grain yield for both –P ( $h^2 = 0.93$ ) and +P ( $h^2 = 0.92$ ) environments. The genetic correlation for grain yield performance between –P and +P conditions was high ( $r_G = 0.89$ ), suggesting that West African sorghum varieties generally possess good adaptation to low-P conditions. However, genotype × phosphorus crossover interaction was observed between some of the highest yielding genotypes from the –P and +P selected sets. Furthermore, direct selection for grain yield in –P conditions was predicted to be 12 % more efficient than indirect selection in +P conditions; and the conclusion was that selection under –P conditions should be useful for sorghum improvement in West Africa.

A final example of using the theory of correlated responses to selection is the paper by Atlin et al. (2001) on breeding cultivars for resource poor farmers in unproductive marginal environments. They use the formula to discuss how such breeding can be made more effective by combining the features of scientist-led, ‘formal’ plant breeding programmes with those of participatory plant breeding. The former is characterized by high heritability in the selection environment and the latter by a high genetic correlation between selection and target environment.

## Days to Flowering and Photoperiod Sensitivity

### *Maize*

Maize (*Zea mays* L. subsp. *mays*) was domesticated from native Mexican teosinte (*Zea mays* L. subsp. *parviglumis*), a species adapted to day lengths of less than 13 h (Hung et al. 2012a, b). Under the longer day lengths of higher latitudes, teosinte flowers very late or not at all. Thus the spread of maize to geographically and ecologically diverse environments, from Canada to Chile, required its adaptation to long day lengths. This was achieved over a long period of time by Meso-American farmers, well before the arrival of Columbus to the Americas. Despite the complexity of flowering time in maize, changes can be made quickly by effective recurrent selection. Teixeira et al. (2015) analyzed the response to selection for early female flowering (silking) in a temperate environment (Iowa USA, 42.03°N) of a population of tropical maize plants. The base population was produced by one generation of open-pollination of five tropical accessions from Brazil, Cuba, Ecuador and Guatemala. Selection was practised for ten generations, with 10,000 individuals grown each generation and 300–500 selected and allowed to open-pollinate in an isolation block. A random sample of plants from remnant seed of the even numbered generations (g0–g10) were self-pollinated, and then random sib-mating was allowed within selfed families to produce enough seed for evaluation. A set of 297 families (55 or 56 from each generation apart from g0 that had only 18) was assessed in nine locations from 43.05°N (Madison, Wisconsin) to 18.00°N (Ponce, Puerto Rico), along with two temperate (relatively photoperiod insensitive B73 and B97) and two tropical (photoperiod sensitive Ki14 and CML254) inbred line controls. Ki14 and CML254 took 103 and 117 days, respectively, to flower in long days but only 87 and 88, respectively, in short days. In contrast, B73 and B97 took 69 and 70 days, respectively, to flower in long days and 77 and 77, respectively, in short days. Thus flowering was delayed by photoperiods longer or shorter than normal for the lines. On average, families from g10 flowered 20 days earlier (23 in long days and 15 in short days) than families from g0, with a 9 day separation between the latest g10 family and the earliest g0 one. The magnitude of the response tended to be greater in the first few generations than in the later generations. The direction of response to selection was consistent in each environment but the greatest responses were at high latitudes, so that some genotype by environment interaction was detected. A strong initial decrease in genotypic variance was followed by a plateau. The authors discuss the results in terms of selection for photoperiod insensitivity, but also point out that the large environmental differences in flowering time between locations can be explained by an increase in minimum daily temperature resulting in earlier flowering, presumably because of faster growth. Locations at higher latitudes have longer days but lower minimum daily temperatures. More general information on the interactions between temperature and photoperiod can be found in the review by Craufurd and Wheeler (2009). The rapid response to selection is an encouraging result when thinking about the rate of adaptation that may be required for appropriate

flowering times and crop durations in response to climate change in a wide range of environments.

The magnitude of genotype by environment interactions for flowering time depends on the environments and germplasm under consideration, as can be seen in the experiments of Hung et al. (2012a, b) which were introduced in Chap. 5. One set of experiments involved a mapping population derived from 25 maize inbreds (tropical and temperate) each crossed to inbred line B73 (temperate), and assessed in 8 long-day length and 3 short-day length environments. Long day lengths delayed flowering time of inbred parents from tropical regions eight times more than inbreds from temperate zones. Flowering time of all families was highly consistent across long-day length environments with a ratio of genotype by environment interaction to genotypic variation of 0.10. In contrast, for families derived from tropical parents crossed to B73 and assessed in long days and short days, the ratio rose to 0.35.

## Rice

Days to flowering and photosensitivity in rice provides an example of both genotype  $\times$  environment interaction and QTL analysis. Maheswaran et al. (2000) derived 143 recombinant inbred lines (RIL) of rice from the cross between CO39 (*indica* rice) and Moroberekan (*japonica* rice), grew replicates of them in a glasshouse in both 10 and 14 h day lengths, and measured days to flowering (DTF or heading date). There was variation between the RILs in the extent to which long days delayed flowering. A standard analysis of variance would partition the variation into one degree of freedom for the average difference (over genotypes) between 10 and 14 h, 142 degrees of freedom for the differences between genotypes averaged over the 2 day lengths, and 142 degrees of freedom for the genotype  $\times$  environment interaction (the variation over genotypes in the difference between 14 and 10 h). The differences between genotypes can be viewed as differences in flowering time *per se* (i.e. days to flowering) and the genotype  $\times$  environment interaction as the photoperiod sensitivity (the extent to which long days delay flowering). The authors also used 127 RFLP markers to map QTLs for days to flowering (DTF) in 10 and 14 h days and for photoperiod sensitivity (PS). They found 15 QTLS for DTF of which four affected PS. The seven QTLs that were detected in both 10 and 14 h days can be viewed as genes for flowering *per se* and the four that affected DFT in one day length and also PS can be viewed as genes for photoperiod sensitivity, of which the one on chromosome 5 had the largest effect. An association was found between QTLs for late flowering and blast resistance, thus providing a genetic explanation for this frequent observation in rice.

A similar experiment done by Xu is described in his book on molecular breeding (Xu 2010). This time doubled haploids (DH) from the cross between Zhaiyeqing 8 and Jingxi 17 were grown in two environments in China (Beijing and Hangzhou) that mainly differ in day length and temperature, with the shorter days and higher

temperatures of Hangzhou favouring earlier flowering. Again days to heading (DTH) were measured and the relative difference  $[(\text{DTH in Beijing} - \text{DTH in Hangzhou}) / (\text{DTH in Beijing})] \times 100$  used as a measure of photo-thermo sensitivity (PTS). Two QTLs on chromosomes 8 and 10 were associated with DTH in both locations, one QTL on chromosome 1 was associated with DTH in Beijing and one on chromosome 12 with DTH in Hangzhou, and one QTL on chromosome 7 was associated with PTS. Thus sometimes separate QTLs can be detected and ascribed functions for overall genetic differences and those associated with genetic responses to environmental change. On other occasions, some QTLs will be detected in all environments and other genuine ones in some environments and not others. The breeder will need to decide whether or not alleles at these loci are worth selecting. More examples can be found in the book by Xu (2010).

# **Chapter 8**

## **Genome Evolution and Polyploidy**

### **Introduction**

It is not my intention in this book to review our current knowledge of genome evolution in plants, but rather to highlight aspects of particular interest and relevance to plant breeders. Part II of this book would not be complete without looking at polyploidy among crop species, but we shall see that this cannot be done without mentioning recent findings on genome evolution, which for me include some surprises. First, however, I am going to present a brief section on molecular evolution followed by one on chloroplast and mitochondrial genomes.

### **Molecular Evolution and Phylogenetic Trees**

At the start of Part II of this book we encountered DNA as the chemical basis of heredity, the revolution in the speed of sequencing DNA brought about by ‘Next-Generation Sequencing’, and the discovery of abundant single nucleotide polymorphisms when genomes were compared. We also encountered the molecular clock of phylogenetics as early as the prologue. How can we use nucleotide polymorphisms to construct phylogenetic trees? An introduction to the subject can be found in the book *Mathematical Models in Biology* by Allman and Rhodes (2004). The following account is based on Chaps. 4 and 5 of their book. It may seem relatively lengthy, but it does explain what we mean by the evolutionary distance between two DNA sequences, from different species, that can be aligned.

## A Markov Model of Base Substitutions

We are going to consider a simple Markov model that makes a lot of assumptions. Let us assume that despite differences, we can align the DNA sequences of a gene from an extant plant and a distant ancestor and examine base substitutions. Let us start with the ancestral sequence ( $S_0$ ). Each site in the sequence is one of four bases  $A$ ,  $G$ ,  $C$  or  $T$  (written in italics in this probability context), chosen randomly according to some probabilities  $P_A$ ,  $P_G$ ,  $P_C$  and  $P_T$  ( $P_A + P_G + P_C + P_T = 1$ ).

We put these four probabilities into a column vector  $\mathbf{p}_0$  so that  $\mathbf{p}_0' = (P_A P_G P_C P_T)$ .

This vector describes the ancestral base distribution, with its entries giving the fraction we would expect to be occupied by each of the four bases.

We model the mutation process over one time step to the  $S_1$  sequence assuming only base substitutions occur so that we do not consider deletions, insertions or inversions. There are 16 conditional probabilities of observing a base substitution which can be put into a  $4 \times 4$  matrix  $\mathbf{M}$ .

$$\mathbf{M} = \begin{vmatrix} |P_{A|A} & P_{A|G} & P_{A|C} & P_{A|T}| \\ |P_{G|A} & P_{G|G} & P_{G|C} & P_{G|T}| \\ |P_{C|A} & P_{C|G} & P_{C|C} & P_{C|T}| \\ |P_{T|A} & P_{T|G} & P_{T|C} & P_{T|T}| \end{vmatrix}$$

Thus  $P_{A|G} = P(S_1 = A | S_0 = G)$ , for example, is the probability of base substitution  $A$  given  $G$  at the ancestral site, and the entries in each row refer to the same descendent base. The entries in each column are all  $\geq 0$  and sum to 1 so that we have a Markov matrix. Let us multiply  $\mathbf{M}$  by  $\mathbf{p}_0$ .

$$\mathbf{M}\mathbf{p}_0 = \begin{vmatrix} |P_{A|A} & P_{A|G} & P_{A|C} & P_{A|T}| & |P_A| \\ |P_{G|A} & P_{G|G} & P_{G|C} & P_{G|T}| & |P_G| \\ |P_{C|A} & P_{C|G} & P_{C|C} & P_{C|T}| & |P_C| \\ |P_{T|A} & P_{T|G} & P_{T|C} & P_{T|T}| & |P_T| \end{vmatrix} = \begin{vmatrix} |P_{A|A}P_A + P_{A|G}P_G + P_{A|C}P_C + P_{A|T}P_T| \\ |P_{G|A}P_A + P_{G|G}P_G + P_{G|C}P_C + P_{G|T}P_T| \\ |P_{C|A}P_A + P_{C|G}P_G + P_{C|C}P_C + P_{C|T}P_T| \\ |P_{T|A}P_A + P_{T|G}P_G + P_{T|C}P_C + P_{T|T}P_T| \end{vmatrix}$$

Now looking at the top row of our new column vector and using the definition of conditional probability [if  $X$  and  $Y$  are two events, the conditional probability of  $Y$  given  $X$  is  $P(Y|X) = P(Y \text{ and } X)/P(X)$ ]:

$P_{A|G}P_G = P(S_1 = A \text{ and } S_0 = G)/P(S_0 = G) \times P(S_0 = G) = P(A \text{ and } G)$  so that

$$\begin{aligned} P_{A|A}P_A + P_{A|G}P_G + P_{A|C}P_C + P_{A|T}P_T &= P(A \text{ and } A) + P(A \text{ and } G) \\ &\quad + P(A \text{ and } C) + P(A \text{ and } T) \end{aligned}$$

$= P(A) = P_A$  because  $A$ ,  $G$ ,  $C$  and  $T$  are mutually exclusive events.

The same applies to the other three rows so that

$$\mathbf{M}\mathbf{p}_0 = \mathbf{p}_1,$$

where  $\mathbf{p}_1$  is the vector of probabilities for the four bases occurring in the sequence  $S_1$ , and  $\mathbf{M}$  is the transition matrix. If the mutation process over the next step is the same and independent of what happened during the first step (past history),  $\mathbf{p}_2 = \mathbf{M}\mathbf{p}_1 = \mathbf{M}\mathbf{M}\mathbf{p}_0$ , and so on over further steps so that after  $t$  steps:

$$\mathbf{p}_t = \mathbf{M}^t \mathbf{p}_0$$

Thus we have a site in a DNA sequence that must be in one of four different states, but may switch from one state to another with time. We assume in our simple model that each site in the sequence behaves identically and independently of every other site so that the probabilities can be determined from the sequence data. In other words, each site is an independent trial of the same probabilistic process.

### ***The Jukes-Cantor Model***

The simplest Markov model of base substitution, the Jukes-Cantor model, assumes that all bases occur with equal probability in the ancestral sequence so that  $\mathbf{p}_0' = (1/4 \ 1/4 \ 1/4 \ 1/4)$ . It also assumes that the conditional probabilities describing an observable base substitution are all the same. Hence if  $1/3\alpha$  is the conditional probability of any particular base substitution occurring, then the 12 off-diagonal entries of  $\mathbf{M}$  will all be  $1/3\alpha$ , and as each column sums to 1, the four diagonal entries will all be  $1 - \alpha$ .

$$\mathbf{M} = \begin{vmatrix} 1 - \alpha & \frac{1}{3}\alpha & \frac{1}{3}\alpha & \frac{1}{3}\alpha \\ \frac{1}{3}\alpha & 1 - \alpha & \frac{1}{3}\alpha & \frac{1}{3}\alpha \\ \frac{1}{3}\alpha & \frac{1}{3}\alpha & 1 - \alpha & \frac{1}{3}\alpha \\ \frac{1}{3}\alpha & \frac{1}{3}\alpha & \frac{1}{3}\alpha & 1 - \alpha \end{vmatrix}$$

As well as being a probability,  $\alpha$  is also the rate at which observable base substitutions occur over one time step, measured in units of (substitutions per site)/(time step). Mutations per site per year are often of the order of magnitude  $1 \times 10^{-9}$ . When mutation rates are constant over a long period of time, there is said to be a molecular clock.

For the Jukes-Cantor model,  $\mathbf{p}_0' = \mathbf{p}_1' = (\frac{1}{4} \ \frac{1}{4} \ \frac{1}{4} \ \frac{1}{4})$ , as determined from  $\mathbf{p}_1 = \mathbf{M}\mathbf{p}_0$ . The vector  $(\frac{1}{4} \ \frac{1}{4} \ \frac{1}{4} \ \frac{1}{4})$  is in fact an eigenvector of  $\mathbf{M}$  with eigenvalue 1. In other words, the base composition of the sequence does not change and can be viewed as an equilibrium base distribution. Hence after  $t$  time steps  $\mathbf{p}_t = \mathbf{M}^t \mathbf{p}_0 = \mathbf{p}_0$ , but what is  $\mathbf{M}^t$ , the 16 conditional probabilities of the various net substitutions over these time steps. For example, what is the probability that a base  $A$  will have mutated to a base  $T$ ?

We need to determine the eigenvectors and eigenvalues of  $\mathbf{M}$  and then make use of theorems from matrix algebra. These can be found in a textbook on matrix algebra such as the one by Searle (1966). The four eigenvectors ( $\mathbf{v}_i$ ) and their eigenvalues ( $\lambda_i$ ) ( $i = 1, 2, 3$  and  $4$ ) are:

$$\begin{aligned}\mathbf{v}_1 &= (1 \quad 1 \quad 1 \quad 1)' & \lambda_1 &= 1 \\ \mathbf{v}_2 &= (1 \quad 1 \quad -1 \quad -1)' & \lambda_2 &= 1 - 1\frac{1}{3}\alpha \\ \mathbf{v}_3 &= (1 \quad -1 \quad 1 \quad -1)' & \lambda_3 &= 1 - 1\frac{1}{3}\alpha \\ \mathbf{v}_4 &= (1 \quad -1 \quad -1 \quad 1)' & \lambda_4 &= 1 - 1\frac{1}{3}\alpha\end{aligned}$$

The columns of  $\mathbf{M}^t$  are  $\mathbf{M}^t(1 \ 0 \ 0 \ 0)', \mathbf{M}^t(0 \ 1 \ 0 \ 0)', \mathbf{M}^t(0 \ 0 \ 1 \ 0)'$  and  $\mathbf{M}^t(0 \ 0 \ 0 \ 1)'$  and these can be written as combinations of the eigenvectors, for example:

$$(1 \ 0 \ 0 \ 0)' = \frac{1}{4}\mathbf{v}_1 + \frac{1}{4}\mathbf{v}_2 + \frac{1}{4}\mathbf{v}_3 + \frac{1}{4}\mathbf{v}_4 \text{ so that}$$

$$\begin{aligned}\mathbf{M}^t(1 \ 0 \ 0 \ 0)' &= \frac{1}{4}\mathbf{M}^t\mathbf{v}_1 + \frac{1}{4}\mathbf{M}^t\mathbf{v}_2 + \frac{1}{4}\mathbf{M}^t\mathbf{v}_3 + \frac{1}{4}\mathbf{M}^t\mathbf{v}_4 \\ &= \frac{1}{4}\lambda_1^t\mathbf{v}_1 + \frac{1}{4}\lambda_2^t\mathbf{v}_2 + \frac{1}{4}\lambda_3^t\mathbf{v}_3 + \frac{1}{4}\lambda_4^t\mathbf{v}_4\end{aligned}$$

Hence  $\mathbf{M}^t =$

$$\begin{vmatrix} \frac{1}{4} + \frac{3}{4}\left(1 - 1\frac{1}{3}\alpha\right)^t & \frac{1}{4} - \frac{1}{4}\left(1 - 1\frac{1}{3}\alpha\right)^t & \frac{1}{4} - \frac{1}{4}\left(1 - 1\frac{1}{3}\alpha\right)^t & \frac{1}{4} - \frac{1}{4}\left(1 - 1\frac{1}{3}\alpha\right)^t \\ \frac{1}{4} - \frac{1}{4}\left(1 - 1\frac{1}{3}\alpha\right)^t & \frac{1}{4} + \frac{3}{4}\left(1 - 1\frac{1}{3}\alpha\right)^t & \frac{1}{4} - \frac{1}{4}\left(1 - 1\frac{1}{3}\alpha\right)^t & \frac{1}{4} - \frac{1}{4}\left(1 - 1\frac{1}{3}\alpha\right)^t \\ \frac{1}{4} - \frac{1}{4}\left(1 - 1\frac{1}{3}\alpha\right)^t & \frac{1}{4} - \frac{1}{4}\left(1 - 1\frac{1}{3}\alpha\right)^t & \frac{1}{4} + \frac{3}{4}\left(1 - 1\frac{1}{3}\alpha\right)^t & \frac{1}{4} - \frac{1}{4}\left(1 - 1\frac{1}{3}\alpha\right)^t \\ \frac{1}{4} - \frac{1}{4}\left(1 - 1\frac{1}{3}\alpha\right)^t & \frac{1}{4} - \frac{1}{4}\left(1 - 1\frac{1}{3}\alpha\right)^t & \frac{1}{4} - \frac{1}{4}\left(1 - 1\frac{1}{3}\alpha\right)^t & \frac{1}{4} + \frac{3}{4}\left(1 - 1\frac{1}{3}\alpha\right)^t \end{vmatrix}$$

The probability, for example, that a site that initially had base  $G$  has base  $A$  after  $t$  time steps is therefore found in column two of row one, namely  $\frac{1}{4} - \frac{1}{4}(1 - 1\frac{1}{3}\alpha)^t$ .

As  $0 < \alpha \leq 1$ ,  $-\frac{1}{3} \leq 1 - 1\frac{1}{3}\alpha < 1$  and hence as  $t \rightarrow \infty$ ,  $(1 - 1\frac{1}{3}\alpha)^t \rightarrow 0$  and all of the elements in  $\mathbf{M}^t \rightarrow \frac{1}{4}$  so that each column is the equilibrium distribution.

More general, and hence more realistic models, can be found in books on molecular evolution such as the one by Yang (2014).

### **Jukes-Cantor Phylogenetic Distance**

The diagonal entries in  $\mathbf{M}^t$  are the conditional probabilities that the base at time  $t$  is the same as the base at time 0. As all of the diagonal entries are equal, at time step  $t$ , the fraction of sites that agree with their initial base is given by the formula:

$$q(t) = \frac{1}{4} + \frac{3}{4} \left(1 - 1\frac{1}{3}\alpha\right)^t$$

and the fraction that are different by the formula:

$$p(t) = 1 - q(t) = \frac{3}{4} - \frac{3}{4} \left(1 - 1\frac{1}{3}\alpha\right)^t$$

At  $t = 0$ ,  $p(t) = 0$  and as  $t \rightarrow \infty$ ,  $p(t) \rightarrow \frac{3}{4}$  because  $(1 - 1\frac{1}{3}\alpha)^t \rightarrow 0$

So as  $t$  increases,  $p(t)$  approaches and never exceeds the value 3/4. If we had the original DNA sequence and a mutated version from some later time we could compare many sites before and after mutation and determine the proportion  $p$  that disagreed. Then for  $0 \leq p \leq 3/4$  we could find a  $t$  with  $p(t) = p$  assuming we know  $\alpha$ . Hence:

$$\begin{aligned} p &= \frac{3}{4} - \frac{3}{4} \left(1 - 1\frac{1}{3}\alpha\right)^t \text{ and} \\ \left(1 - 1\frac{1}{3}\alpha\right)^t &= 1 - 1\frac{1}{3}p \text{ so that} \\ t &= \ln\left(1 - 1\frac{1}{3}p\right) / \ln\left(1 - 1\frac{1}{3}\alpha\right) \end{aligned}$$

Although we cannot estimate  $t$  without knowing  $\alpha$ , the product of  $t$  (the number of time steps) and  $\alpha$  (number of substitutions per site per time step) is more meaningful for our purpose as it measures the expected number of substitutions per site during the elapsed time. This expected number of substitutions includes those that we do not observe because they are hidden by subsequent substitutions. With a small time step,  $\alpha$  will be close to zero so that  $\ln(1 - 1\frac{1}{3}\alpha)$  will be very close to  $(-1\frac{1}{3}\alpha)$ . Hence:

$$t \approx -3 \left[ \ln\left(1 - 1\frac{1}{3}p\right) \right] / 4\alpha$$

and the Jukes-Cantor distance between two DNA sequences is:

$$d_{JC} = t\alpha = -\frac{3}{4} \ln\left(1 - 1\frac{1}{3}p\right)$$

If the Jukes-Cantor model accurately describes the evolution of one sequence into another, it estimates the total number of substitutions per site that occurred during the evolution. If the molecular clock hypothesis (constant mutation rate over evolutionary history) is valid, the distance is proportional to the amount of elapsed time, with the constant of proportionality the mutation rate. If there is an independent estimate of time (geological record), then the mutation rate can be found from  $d_{JC}$ . It can be shown that Jukes-Cantor distances are additive and symmetrical. It does not matter which sequence is the ancestral one for calculating the distance. Hence we can calculate the distance between two descendent sequences and construct a phylogenetic tree.

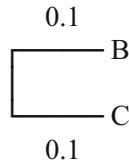
## ***Phylogenetic Trees***

Phylogenetic trees attempt to relate and show the evolutionary history of taxa from a common ancestor at the root of the tree. Here we are concerned with trees constructed from DNA sequence data where each taxon under study has provided a DNA sequence from a gene (or segment of DNA) common to all taxa. The branching structure of a tree is its topology and the tree is bifurcating if three edges (line segments) meet at each interior vertex and two edges meet at the root. The end of an edge at a taxon is called a terminal vertex or leaf. The number of different topological trees that might relate several terminal taxa increases extremely rapidly with the number of taxa. The lengths of edges in a metric tree constructed from DNA sequence data represent the amount of mutation that occurred between splittings of the lineage. If the Jukes-Cantor model of base substitution adequately describes the evolution of several taxa, then the edge length might be the Jukes-Cantor distance between the sequences at the two ends of the edge. The total distance between two taxa along a tree should then be the Jukes-Cantor distance between them. If the molecular clock assumption holds, the amount of mutation along any edge is proportional to the elapsed time. Furthermore, for a rooted metric tree, every leaf will be located the same total distance from the root because every taxon has had the same amount of time to evolve from the root ancestor. For increased readability, vertical lines that contribute nothing to the amount of mutation are often added to a tree. All of these points should be clearer from the very simple example that follows.

Let us assume that we have aligned orthologous DNA sequences from four extant species, A, B, C and D. We therefore have information on terminal taxa from which to construct a phylogenetic tree, but no information on their ancestors. In this very simple example we are going to use the observed  $p$  values and their corresponding Jukes-Cantor distances to build a tree that accounts for the total distances between terminal taxa along the tree.

	Observed $p$ values				Corresponding $d_{JC}$ distances			
	A	B	C	D	A	B	C	D
A		0.245	0.245	0.295		0.4	0.4	0.5
B			0.135	0.295			0.2	0.5
C				0.295				0.5

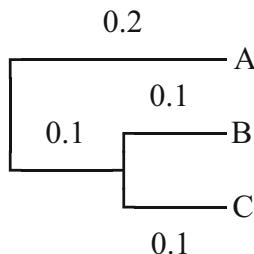
We are going to use the average distance method known as the unweighted pair-group method with arithmetic means (UPGMA). The method produces a rooted tree and assumes a molecular clock. We choose the two closest taxa, B and C, which are 0.2 apart. We draw the following figure with each edge 0.1 (0.2/2) apart. The vertical line is only there for clarity; it has no meaning.



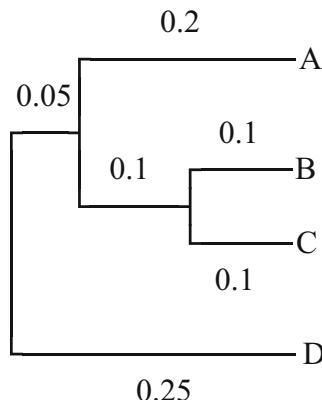
We now combine B and C into a group and average the distances of B and C to each different taxon to get the distance from the group to that taxon.

$B - C$	$A$	$D$
$B - C$	$(0.4 + 0.4)/2 = 0.4$	$(0.5 + 0.5)/2 = 0.5$
$A$		$0.5$

The closest taxa are now group B-C and A, which are 0.4 apart, and we draw the following figure in which the distances A to B and A to C are both 0.4 so that the distance of A to their common ancestor is 0.2.



Again combining taxa, we form a group A-B-C, and compute its distance from D by averaging the original distances of D to each of A, B and C:  $(0.5 + 0.5 + 0.5)/3 = 0.5$ . As this is our last taxa, we know that D is 0.25 (0.5/2) from the root of the tree.



If we had more taxa to relate, we would need to do more steps, but there would be no new ideas involved apart from making sure that the distance between two groups is the average of all the distances from members of one group to another. In practice, a tree will not fit the data exactly as happened here with distances chosen to provide a simple example. Furthermore, a tree will be fitted by computer software based on an appropriate algorithm. Some methods require the presence of a more distantly related ‘outgroup’ to ensure a rooted tree with a common ancestor. More sophisticated methods are required when one cannot assume a molecular clock. The ones likely to be used are those based on neighbour joining or optimality criteria such maximum likelihood or maximum parsimony, a rather different approach that searches for the tree that can be explained with the fewest mutations. Details and arguments about the strengths and weaknesses of various methods can be found in books on molecular evolution such as the one already mentioned by Yang (2014). Plant breeders will commonly encounter and need to interpret phylogenetic trees in the scientific literature.

## **Chloroplast and Mitochondrial Genomes**

It is generally accepted that mitochondria and plastids evolved from free-living prokaryotes. During a long process of endosymbiosis most of their genes were transferred to the nucleus of the host eukaryotic cell which they originally invaded. Hence most of the proteins that they need are now encoded in the nucleus, synthesized in the cytosol, and imported into the organelles. As a consequence of numerous mitochondria and plastids in a cell, numerous copies of their genomes are present. A mixed cell (heteroplasmic) contains more than one allele at a genetic locus. Somatic segregation (sorting out) during the production of daughter organelles and cells can result in a mixture of homoplasmic cells. The mechanisms controlling replication and distribution of organelle-DNA molecules at cell division

are more flexible than the stringent nuclear one, and hence the organelles are said to have a relaxed genome. In cultivated plants, organelles usually display mono-parental inheritance (i.e. inherited from only one parent).

The plastid genome (plastome) of higher plants is a unicircular DNA molecule of about 150 kbp. It encodes about 100–150 genes, including most of the genes involved in photosynthesis as well as those for its genetic machinery (rRNAs, tRNAs and ribosomal proteins). One gene of particular interest for crop production encodes the large sub-unit of ribulose 1,5-biphosphate carboxylase/oxygenase (RuBisCO), the key enzyme in photosynthesis and photorespiration.

The mitochondrial genome (chondriome) of higher plants comprises a population of circular and linear DNA molecules with different sizes and sequences. Internal recombination events can result in new arrangements so that gene order is not as strict as in the plastome. The size of the chondriome is also more variable among plant species, ranging from 200 to 2000 kbp, but does not correlate with the number of genes. The chondriome encodes a number of proteins involved in the electron transfer chain and the ATPase complex, but its genetic machinery is incompletely encoded.

## Classification of Polyploids

Classical cytogenetics (1900–1953) revealed that many cultivated plants are polyploids, having more than the two sets of chromosomes found in the fertilized eggs of diploid species, where one (haploid) set comes from the pollen and the other homologous set from the unfertilized egg. We shall encounter polyploid species throughout the rest of this book. Autopolyploids have more than two sets of chromosomes of the same species. The potato, for example, is an autotetraploid ( $2n = 4x = 48$ ) that we shall shortly examine in more detail. Autopolyploids arise either from a failure of mitosis or a failure of meiosis leading to non-reduction (unreduced gametes). Allopolyploids have chromosome sets from more than one species (A, B, C, D, etc.), two sets of each species being required for normal fertility. *Brassica napus* (oilseed rape) (AACC,  $2n = 38$ ), for example, is the allotetraploid of *B. rapa* (AA,  $2n = 2x = 20$ ) and *B. oleracea* (CC,  $2n = 2x = 18$ ) that we shall also shortly examine in more detail. As mentioned in Chap. 2, plant breeders developed methods to increase the natural low frequency of chromosome doubling, particularly the use of the chemical colchicine from 1937 (Blakeslee and Avery 1937). As a result, well established polyploid species were re-synthesized to widen their genetic base for plant breeding, and new auto- and allopolyploids were created. Later in this book we shall meet autotetraploid cultivars of red clover (*Trifolium pratense*) produced by colchicine treatment of diploids, and shortly we shall consider triticale, the hybrid between wheat and rye. In extreme autopolyploidy the genomes are structurally identical, with unrestricted recombination, so that they are functionally polyploid and display polysomic inheritance. In extreme allopolyploidy the genomes are so differentiated that pairing and recombination are

restricted to homologous chromosomes, and the allopolyploid is functionally diploid and displays disomic inheritance (amphidiploid). Between these extremes, a complete range of intermediate types is possible, including segmental allopolyploids (Grant 1971). The latter have subgenomes (e.g. A<sub>s</sub> and A<sub>t</sub>) that are differentiated structurally to an intermediate degree so that there is some pairing between them in the polyploid (e.g. A<sub>s</sub>A<sub>s</sub>A<sub>t</sub>A<sub>t</sub>). We are now going to look at the production of unreduced gametes in diploids and meiosis and the phenomenon of double reduction in autotetraploids, before considering some examples of polyploids. Chromosome behaviour and inheritance patterns are even more complex in higher levels of autoploidy, but details will not be given in this book.

## Unreduced Gametes from Diploid Species

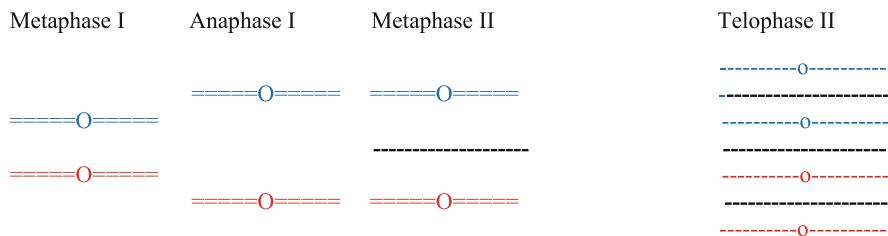
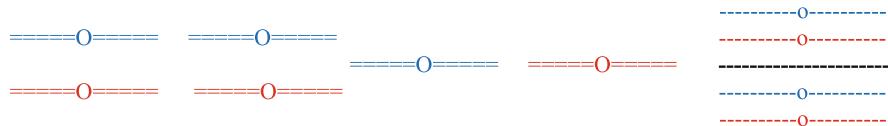
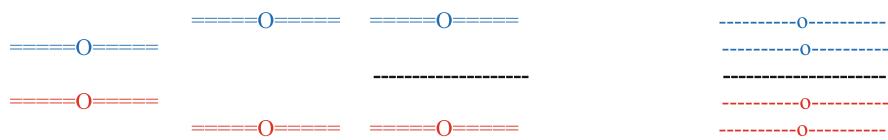
### *Potatoes*

Let us return to the domestication of potato, which was introduced in Chap. 1. The result of domestication was a diploid cultigen *S. tuberosum* Group Stenotomum (Dodds 1962) from which all the other cultivated potatoes of South America were derived. Group Stenotomum potatoes remained confined to the central Andes of Peru and Bolivia. In contrast, their tetraploid (four instead of two sets of chromosomes) descendants became grown throughout the upland Andes of South America (Group Andigena), and a secondary derivative of these became adapted to the long days of coastal Chile (Group Tuberorum). The outcome was a large number of tetraploid landraces adapted to a wide range of growing conditions; but how did Group Andigena arise from Group Stenotomum? Sukhotu and Hosaka (2006) concluded from chloroplast and nuclear DNA markers that Group Andigena arose from Group Stenotomum through sexual polyploidization from unreduced gametes many times at many places in the fields of Group Stenotomum. These tetraploids were subsequently modified by occasional and unintentional selection of natural hybrids with neighbouring wild species to give present-day Group Andigena. Scurrah et al. (2008) have shown that closely related species growing around farmers' fields can hybridize with Group Andigena and that some hybrid progeny would be selected by present day Andean farmers. These results explain the chromosome behaviour and tetrasomic inheritance of tetraploid *S. tuberosum* and why it can be regarded as the autotetraploid of diploid Group Stenotomum for practical purposes.

## ***Unreduced Gametes and Sexual Polyploidization***

A number of mechanisms during the premeiotic, meiotic and postmeiotic stages of gamete formation may produce  $2n$  gametes (Veilleux 1985), referred to as diplandroid and diplogynoid for pollen and egg, respectively. They can be viewed as abnormal events resulting from genetic mutations that affect the outcome of meiosis. In potatoes they can be classified genetically into first division restitution (FDR) or second division restitution (SDR) (Peloquin et al. 1999), as shown in Fig. 8.1. Starting with microsporogenesis, FDR can arise from an abnormal parallel orientation of the spindles during metaphase II and anaphase II that prevents cell division, and consequently two  $2n$  microspores are formed following cytokinesis. The trait is simply inherited as a Mendelian recessive (*ps*), as reported by Mok and Peloquin (1975). FDR can also arise from fused spindles during the second meiotic division, but the genetic consequences are the same (Ramanna 1979). In contrast, SDR occurs when the first division of meiosis is followed by a premature cytokinesis so that the second division does not occur, but the chromatids do separate. Consequently a dyad of two  $2n$  microspores is formed. Premature cytokinesis is also simply inherited as a Mendelian recessive (*pc-1* and *pc-2*) (Mok and Peloquin 1975). Synaptic mutants have been found which cause poor pairing and/or reduced chiasmata frequencies in microsporogenesis (Jongedijk and Ramanna 1988). When the synaptic mutant *sy-3* is combined with ‘parallel spindles’, it can produce FDR  $2n$  pollen with no crossing over (NCO), and hence the mechanism is called FDR-NCO. In megasporogenesis, the predominant mechanism of  $2n$  egg formation is omission of the second meiotic division, again simply inherited as a Mendelian recessive (*os*), and resulting in two  $2n$  megasporules of which one is functional (Werner and Peloquin 1990). Another simply inherited mechanism is failure of cytokinesis after the second meiotic division (*fc*). Both mechanisms are genetically second division restitutions. FDR  $2n$  eggs can be produced through simply inherited desynapsis (*ds-1*) resulting in a direct equational division of univalent chromosomes at anaphase I (pseudohomotypic division) (Jongedijk et al. 1991).

The significance of  $2n$  gametes in potato breeding stems from the fact that most of the wild relatives of our principal cultivated potato (*Solanum tuberosum*) are diploid species. Furthermore, since 1958 it has been possible to produce relatively large numbers of dihaploids (diploids) from tetraploid *Solanum tuberosum* (Hougas et al. 1958). Diploid species and dihaploid-species hybrids in general produce both  $2n$  and  $n$  gametes (Mendiburu and Peloquin 1977). Matings between  $4x$  and  $2x$  parents ( $4x\text{-}2x$  and  $2x\text{-}4x$  crosses) give rise to almost entirely  $4x$  progeny due to a ‘triploid block’ mechanism. Matings between  $2x$  parents ( $2x\text{-}2x$  crosses) produce  $2x$  and  $4x$  offspring, the frequencies being cross dependent (Mendiburu and Peloquin 1977). Hence the  $4x$  offspring have to be distinguished by chromosome counts or a ploidy correlated trait such as mean number of chloroplasts in stomata (Tai 1994). In contrast, in bananas and citrus species, ploidy manipulation has been used to produce seedless triploid cultivars, as we shall see in the examples of polyploids.

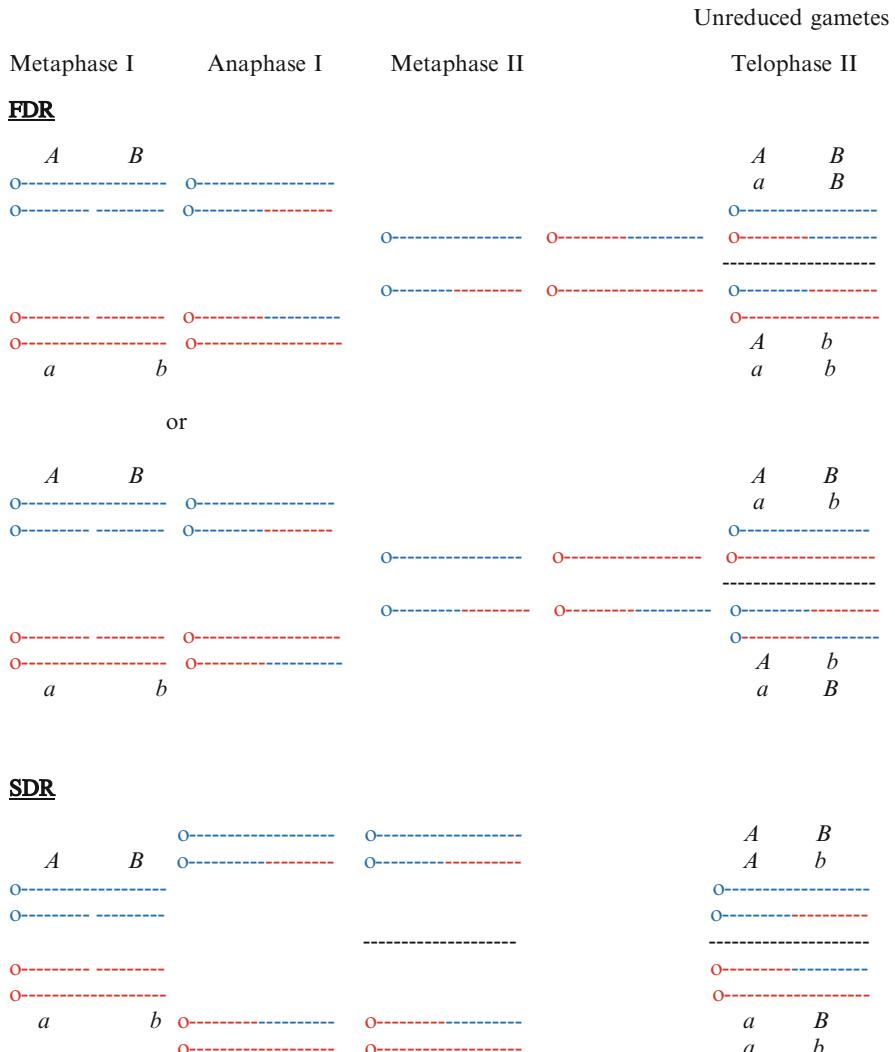
**Normal****FDR****SDR**

**Fig. 8.1** Reduced gametes from normal meiosis in diploids and unreduced gametes from first division restitution (FDR) and second division restitution (SDR): diagrams show two homologous chromosomes, centromeres (O and  $\textcircled{O}$ ) and adjoining sections of chromosomes (chromatids ----- and ---) but no chiasmata, and cytokinesis (----)

## The Genetic Consequences of FDR and SDR in Diploids

We will now briefly consider the genetic consequences of FDR and SDR. More details can be found in the reviews by Tai (1994) and Peloquin et al. (1999). We need to return to the diagram above, add a chiasma, and look at the consequences for heterozygous loci between the centromere and chiasma ( $A/a$ ) and beyond the chiasma ( $B/b$ ) (Fig. 8.2).

With FDR, no-exchange tetrads (NET) are formed (strictly dyads are formed) when there is no crossing over between chromatids carrying  $A$  and  $a$ . All  $2n$  gametes are of genotype  $Aa$ . Single-exchange tetrads (SET) are formed when there is crossing over between chromatids carrying  $B$  and  $b$ . Two types of



**Fig. 8.2** The genetical consequences of producing unreduced gametes by FDR and SDR

chromosomal orientation are possible at metaphase II with SET, as shown in Fig. 8.2. The first orientation results in  $2n$  gametes with genotypes  $BB$  and  $bb$  in a 1:1 ratio, whereas the second results in all  $Bb$  genotypes. If the two orientations are equally likely, then we expect ratios of  $1/4BB:1/2Bb:1/4bb$ . If the frequency of SET is  $\beta$ , then the ratio of  $BB, Bb$  and  $bb$  gametes is  $1/4\beta:(1 - 1/2\beta):1/4\beta$ . If one chiasma is equally likely to occur anywhere along a chromosome arm, the chromosome is 50 cM long and  $\beta$  increases from zero to one in a linear fashion as you move from

centromere to the end of the chromosome arm. Furthermore the average value of  $\beta$  for the chromosome arm is one half, and the average frequency of transmission of heterozygosity from diploid parent to  $2n$  gamete is 0.75. If we extend considerations to two-, three- and four-strand double crossovers (DET), then for a heterozygous locus ( $C/c$ ) beyond the second chiasma we find ratios of  $1/8CC:3/4Cc:1/8cc$ ; in other words a heterozygous frequency of 0.75, as shown by Cuenca et al. (2011). With two crossovers per chromosome arm, we have an arm length of 100 cM (one Morgan) and a problem. What is the distribution of these two crossovers? Do they occur at random or do they interfere with one another? Cuenca et al. (2011) considered a number of possibilities including interference. Here I am just going to look at one simple situation (Table 8.1). The level of heterozygosity falls from 100 % at the centromere to 75 % one third the way along, to 66.7 % two thirds the way along, and back to 75 % at the end of the chromosome arm; with an average value of near 75 % for the whole arm.

With SDR and NET, the  $2n$  gametes have genotypes  $AA$  and  $aa$  with equal frequency, whereas with SET, the chromosomal orientation at metaphase II does not alter the outcome of only  $Bb$  gametes. Again assuming the frequency of SET is  $\beta$ , the ratio of  $BB$ ,  $Bb$  and  $bb$  gametes is  $1/2(1 - \beta):\beta:1/2(1 - \beta)$ . If one chiasma is equally likely to occur anywhere along a chromosome arm,  $\beta = 1/2$ , and the frequency of transmission of heterozygosity from diploid parent to  $2n$  gamete is 0.50. When  $\beta = 2/3$ , the ratio of  $BB$ ,  $Bb$  and  $bb$  gametes is 1:4:1 for both FDR and SDR. If we extend considerations to two-, three- and four-strand double crossovers (DET) for SDR, then for a heterozygous locus ( $C/c$ ) beyond the second chiasma we find ratios of  $1/4CC:1/2Cc:1/4cc$ ; in other words a heterozygous frequency of 0.50, as shown by Cuenca et al. (2011). Again with two crossovers per chromosome arm, I am just going to consider the same situation as for FDR (Table 8.1). The level of heterozygosity increases from 0 % at the centromere to 50 % one third the way along, to 66.7 % two thirds the way along, and back to 50 % at the end of the chromosome arm; with an average value of near 50 % for the whole arm. Different answers are obtained with different models (Cuenca et al. 2011), but in all cases the average frequency of transmission of heterozygosity from diploid parent to  $2n$  gamete is much higher for FDR (about 75 %) than for SDR (about 50 %). Furthermore, from 0 cM at the centromere to 33 cM, it is less than 50 % with SDR and more than 75 % with FDR. FDR gametes also maintain and transmit a large amount of inter-locus epistatic interactions, which are important in maximizing heterosis, but are not considered here.

Tai (1994) reviewed the genetic consequences of  $2n$  gametes for sexual polyploidization in  $2x \times 2x$  crosses (FDR  $\times$  FDR, FDR  $\times$  SDR and SDR  $\times$  SDR) and  $4x \times 2x$  crosses ( $4x \times$  FDR and  $4x \times$  SDR), where the tetraploid genotype can have up to four different alleles ( $A_1A_2A_3A_4$ ) at a locus. He concluded that the overall advantage of FDR in transmitting heterozygosity from  $2x$  parents to  $4x$  progeny depends on the allelic diversity present in the parents. In diploid by diploid crosses, the SDR  $\times$  SDR cross is expected to be inferior for maintaining hybrid vigour in the

**Table 8.1** Expected heterozygosity ( $H\%$ ) of locus  $l$ , which is  $d$  cM from centromere on chromosome arm of length 100 cM, in unreduced gametes produced by FDR and SDR; assuming that there are always two chiasmata ( $\chi$ ) distributed at random along chromosome arm where  $P$  is the probability of a chiasma between the centromere and locus  $l$

progenies unless the diallelic genotypes have the same degree of heterosis as the tri- and tetra-allelic ones. Furthermore, the FDR  $\times$  FDR cross is favoured for the possibility of creating a genetically more uniform hybrid population than the other two types of crosses, since it has a higher frequency of tetra-allelic genotypes ( $A_1A_2A_3A_4$ ) for loci up to 30 cM (out of 50 cM) from the centromere, and a low frequency of diallelic genotypes for loci located anywhere in a chromosome. The use of  $2n$  gametes in potato breeding has been reviewed by Tai (1994) and Ortiz (1998), and more recently by Carpoto and Frusciante (2011). We shall return to this topic in Chap. 15.

## Tetrasomic Inheritance in Autotetraploids

### *Meiosis in Autotetraploids*

In a true autotetraploid the four sets of chromosomes are entirely homologous, meiotic pairing within each group of four homologous chromosomes is completely random and tetrasomic inheritance always occurs. The first meiotic division begins with competition between homologous chromosomes as they try to form pairs, followed by chiasma formation and crossing over. With four chromosomes there are a number of possible outcomes depending upon whether or not changes of pairing partner occur along the lengths of the chromosomes. In the simplest situation the four homologous chromosomes pair at random to give two bivalents (2II's) and the two divisions of meiosis proceed normally to give equally viable diploid gametes. Male and female gametes then unite at random to form tetraploid offspring. More complex pairing patterns can result in quadrivalents or various combinations of trivalents, bivalents and univalents. Hence quadrivalent formation is evidence for autotetraploidy, but its absence does not prove disomic inheritance. Quadrivalents can be a chain (CIV) or a ring (RIV), with three chiasmata sufficient for a chain and four for a ring. Five or more chiasmata would give a complex pattern but may not occur in practice, particularly with short chromosomes. Quadrivalent orientations on the spindle that result in 2 + 2 disjunction (i.e. the migration of two chromosomes to each pole) produce balanced diploid chromosome sets. Where a trivalent and univalent (CIII + I) occur, the trivalent will usually disjoin two and one, and the univalent to one or other pole, or be left behind and lost. The simplest trivalent has two chiasmata and forms a chain, but more complex arrangements are possible, particularly if there are three or more chiasmata. One bivalent and two univalents (II + 2I's) and four univalents (4I's) are the remaining possibilities. The second division of meiosis is straightforward in that the centromeres lie on the spindle equator and then split into daughter centromeres which disjoin to opposite spindle poles. Overall, some but not all outcomes of meiosis in autotetraploids lead to balanced diploid chromosome sets and hence viable and fertile diploid gametes. Thus some reduction in fertility must be anticipated, but may or may not occur. In

potatoes low frequencies of quadrivalents, trivalents and univalents do occur, but bivalents predominate (Swaminathan and Howard 1953). In alfalfa (lucerne), Bingham and McCoy (1988) found that the full complement of 16 bivalents at metaphase I was normal.

## ***Double Reduction in Autotetraploids***

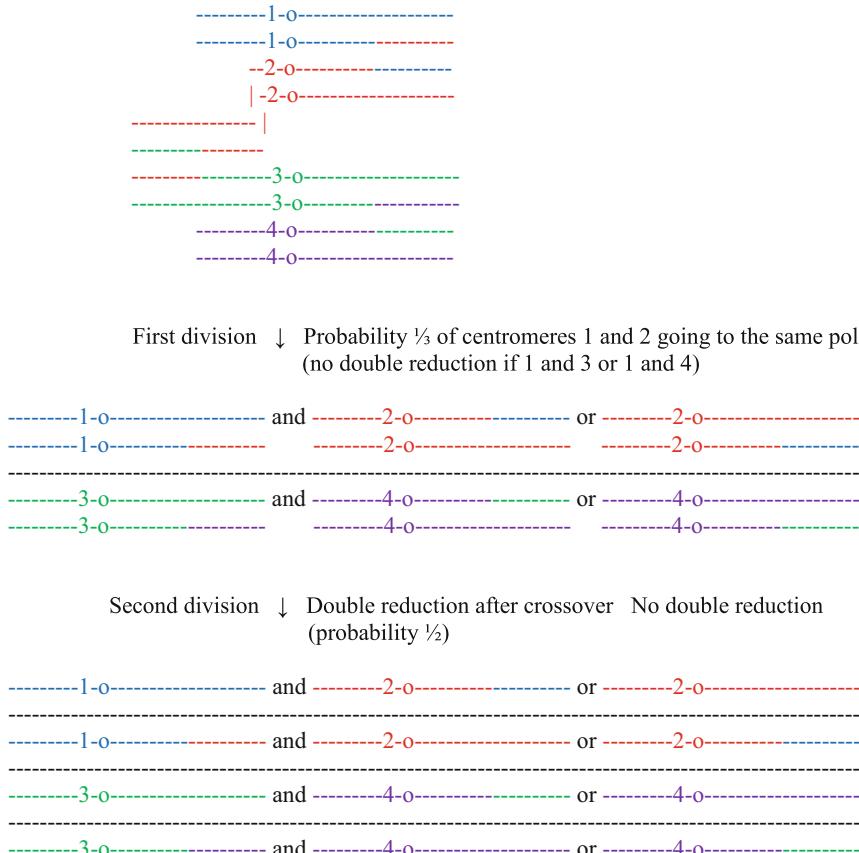
In autotetraploids, sister chromatids can finish in the same (diploid) gamete, a phenomenon known as double reduction (Mather 1936). This happens when four homologous chromosomes form a quadrivalent, two of the chromosomes with a crossover go to the same pole at first anaphase so that beyond the crossover sister chromatids are attached to different centromeres, and finally at second anaphase the two sister chromatids go to the same pole. The coefficient of double reduction ( $\alpha$ ) is defined as the probability of two sister chromatids going to the same gamete (Fisher and Mather 1943). In the example just considered and illustrated in Fig. 8.3, the maximum value for the coefficient of double reduction is 1/6. However, if a double crossover is possible between the centromere and the section of chromatid under consideration, the maximum coefficient is 1/4, as pointed out by Luo et al. (2006).

The gametic output of an autotetraploid follows from the definition of  $\alpha$ . Each of the four homologous chromosomes ( $A_1A_2A_3A_4$ ) can provide a pair of sister chromatids so that the probability for any one chromosome is  $1/4\alpha$ . Then for the  $(1 - \alpha)$  situations where this does not occur, two chromosomes (non sister chromatids) are chosen at random so that the overall gametic outcome from ( $A_1A_2A_3A_4$ ) is:

$$\frac{1}{4}(A_1A_1 + A_2A_2 + A_3A_3 + A_4A_4)\alpha + \frac{1}{6}(A_1A_2 + A_1A_3 + A_1A_4 + A_2A_3 + A_2A_4 + A_3A_4)(1 - \alpha)$$

When the coefficient is zero, one has chromosomal segregation whereas when it is 1/7 one appears to have random chromatid segregation; but meiosis is always a two division process with ‘half-centromeres’ and chromatids separating in the second division.

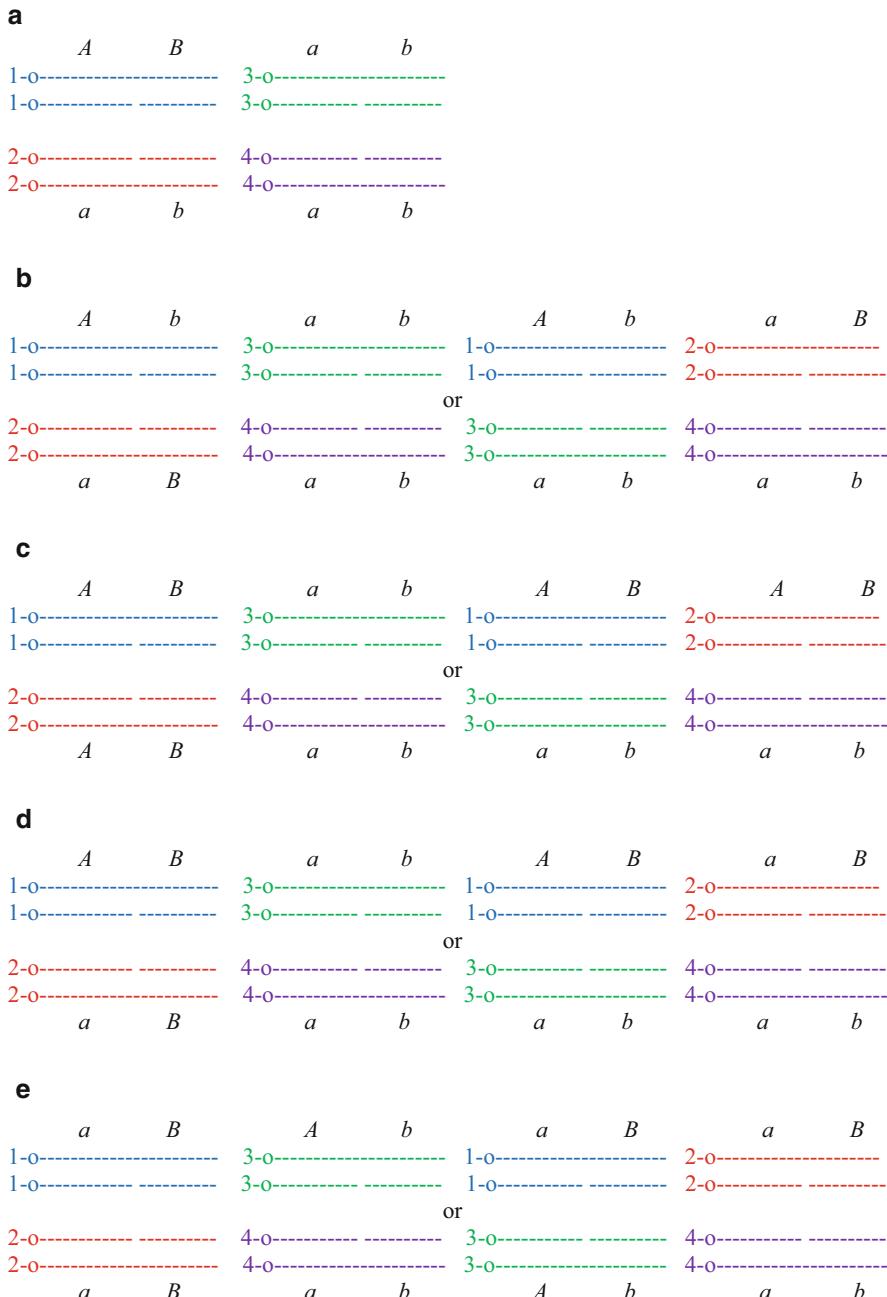
We are now going to look at the theory and practice of linkage analysis in autotetraploids. For simplicity, we shall assume chromosomal segregation: the four homologous chromosomes pair at random to give two bivalents, the two divisions of meiosis proceed normally and give equally viable diploid gametes, and male and female gametes then unite at random to form tetraploid offspring.



**Fig. 8.3** Double reduction in an autotetraploid (frequency 0 before and  $1/6$  after crossover). Four homologous chromosomes pair: 1 and 4 plus 2 and 3 before the centromeres; 1 and 2 plus 3 and 4 after the centromeres, with crossing over between pairs of chromatids (sister chromatids are shown in same colour)

### Linkage Analysis with Chromosomal Segregation

Let us start by considering two simplex loci ( $Aaaa$  and  $Bbbb$ ) with dominant alleles  $A$  and  $B$  in coupling ( $AB/ab/ab/ab$ ) and repulsion linkage ( $Ab/aB/ab/ab$ ); the recombination frequency being  $r$ . We have to consider all of the possibilities that can occur during meiosis (not all are shown in Fig. 8.4). The three possible chromosome pairings are 12 + 34, 13 + 24 and 14 + 23. With pairing 12 + 34 centromeres 1 and 3 or 1 and 4 go to the same pole at anaphase I and then the centromeres divide and go to opposite poles at anaphase II, there having been two possible spindle orientations at metaphase II. With simplex/simplex coupling all three pairings give the same result (Fig. 8.4a), whereas with simplex/simplex repulsion only 12 + 34 (Fig. 8.4b) is informative about recombination.



**Fig. 8.4** Linkage analysis in autotetraploids with chromosomal segregation (two bivalents). (a) Simplex/simplex coupling. (b) Simplex/simplex repulsion. (c) Duplex/duplex coupling. (d) Simplex/duplex coupling. (e) Simplex/duplex repulsion

The outcome of a backcross to  $ab/ab/ab/ab$  is as follows.

Simplex/simplex coupling

$AB$		$ab$
$ab$	$\times$	$ab$
$ab$		$ab$
$ab$	$\downarrow$	$ab$

Expected genotypes

$AB$	$Ab$	$aB$	$ab$
$ab$	$ab$	$ab$	$ab$
$ab$	$ab$	$ab$	$ab$
$ab$	$ab$	$ab$	$ab$
$\frac{1}{2}(1-r)$	$\frac{1}{2}(r)$	$\frac{1}{2}(r)$	$\frac{1}{2}(1-r)$

Simplex/simplex repulsion

$Ab$		$ab$
$aB$	$\times$	$ab$
$ab$		$ab$
$ab$	$\downarrow$	$ab$

Expected genotypes from 12 + 34

$AB$	$Ab$	$aB$	$ab$
$ab$	$ab$	$ab$	$ab$
$ab$	$ab$	$ab$	$ab$
$ab$	$ab$	$ab$	$ab$
$\frac{1}{2}(r)$	$\frac{1}{2}(1-r)$	$\frac{1}{2}(1-r)$	$\frac{1}{2}(r)$

Expected phenotypes from 13 + 24 or 14 + 23

$$\frac{1}{4}AB \quad \frac{1}{4}Ab \quad \frac{1}{4}aB \quad \frac{1}{4}ab$$

Hence the four phenotypes occur with the following frequencies:

Phenotypes	AB	Ab	aB	ab	Total
Coupling	$(1/2)(1-r)$	$(1/2)(r)$	$(1/2)(r)$	$(1/2)(1-r)$	1
Repulsion	$(1/6)(1+r)$	$(1/6)(2-r)$	$(1/6)(2-r)$	$(1/6)(1+r)$	1
Observed numbers	$a$	$b$	$c$	$d$	$n$

The maximum likelihood estimate of  $r$  and its standard error is (Hackett et al. 1998):

$$r = (b+c)/n, \quad SE = [r(1-r)/n]^{1/2} \quad \text{for simplex/simplex coupling and}$$

$$r = [2(a+d) - (b+c)]/n, \quad SE = [(1+r)(2-r)/n]^{1/2} \quad \text{for simplex/simplex repulsion.}$$

If  $r = 0.10$ , and  $n = 100$ , its SE with coupling linkage is 0.03, but 0.14 with repulsion linkage.

These results demonstrate that it is difficult to detect repulsion linkages with tetrasomic inheritance. With simplex markers one is therefore likely to produce a linkage map comprising linkage groups equal in number to that of the tetraploid chromosomes ( $4x$ ), having failed to identify homologous chromosomes. However, the situation improves when we have dominant duplex markers and consider duplex/duplex coupling (Fig. 8.4c) and simplex/duplex coupling (Fig. 8.4d) and repulsion (Fig. 8.4e) linkages. With these linkages the  $12+34$  pairing is not informative whereas the other two possible pairings are informative about recombination. Estimation of  $r$  and its standard error can be found in the paper by Hackett et al. (1998). In the absence of explicit formulae for the recombination fractions, the log-likelihood is maximized numerically.

The standard errors for  $r = 0.10$ , and  $n = 100$  are as follows.

$$\begin{aligned} \text{SE} &= [3r(2-r)(5-2r+r^2)]^{1/2}/[4n(5-6r+3r^2)]^{1/2} && \text{for duplex/duplex coupling} = 0.039 \\ \text{SE} &= [r(1-r)(2+r)(3-r)]^{1/2}/[n(1+r-r^2)]^{1/2} && \text{for simplex/duplex coupling} = 0.071 \\ \text{SE} &= [r(1-r)(2+r)(3-r)]^{1/2}/[n(1+r-r^2)]^{1/2} && \text{for simplex/duplex repulsion} = 0.071 \end{aligned}$$

The graphs of SE against  $r$  (for  $r = 0\text{--}0.5$ ) can be found in the paper by Hackett et al. (1998). For  $r$  between 0 and 0.1 the SE for duplex/duplex coupling is only slightly larger than that for simplex/simplex coupling, and the SE for simplex/duplex linkage (either coupling or repulsion) is at worst half that for simplex/simplex repulsion. Hackett et al. (1998) went on to show by computer simulation that these combinations of markers can be used to identify and map four homologous chromosomes in an autotetraploid. They concluded that a population size of at least 150 individuals should be used, and that a larger number, say 250, would provide a better chance of identifying homologous chromosomes. Since 1998 further progress has been made in the theory and practice of linkage analysis in autotetraploids. The best tetraploid map of potatoes published to date is the one by Hackett et al. (2013) in which they extended the methodology to use dosage data on SNPs. The two parents (processing clone 12601ab1 and table cultivar Stirling) and 190 F<sub>1</sub> offspring were genotyped using an Infinium 8300 Potato SNP Array. Overall 3839 of the 5378 polymorphic SNPs were assigned putative genetic locations. Between 74 and 152 SNPs were mapped to each of the 12 linkage groups (chromosomes) and between 100 and 300 further SNPs were allocated to approximate bins. The length of chromosomes, calculated using the mapped SNPs, varied from 71.9 to 121.6 cM, with a total length of 1087.5 cM. A low level of double reduction products was detected: on average 2.7 out of 190 individuals for Stirling as parent and 3.4 for 12601ab1. The SNP panel was selected to target single copy regions of the genome. Hence for the majority of SNPs it was possible to assign a unique genomic location in the published potato genome sequence. This direct ‘map to genome’ link should facilitate the identification of candidate genes at trait loci.

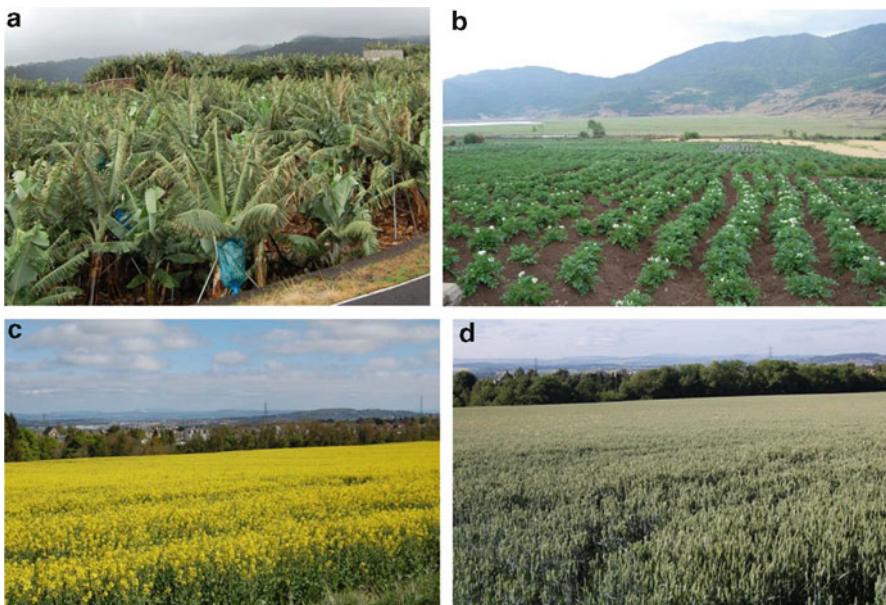
## *QTL Analysis in Autotetraploids*

QTL analysis has developed alongside that for linkage analysis of autotetraploids with chromosomal segregation. The most reliable analysis of the phenotypic data from the cross between potato processing clone 12601ab1 and table cultivar Stirling is the one by

Hackett et al. (2014) using the linkage map described in the previous section (Hackett et al. 2013). These papers extend the theory and application of the interval mapping method of Hackett et al. (2001) to the use of dosage data on SNPs. For each offspring, Hackett et al. (2001) used the marker information on each chromosome to identify possible configurations of chromosomes inherited from the two parents and the locations of crossovers on these chromosomes. A branch and bound algorithm identified configurations with the minimum number of crossovers. From these configurations, the conditional probability of each possible QTL genotype was estimated for a series of positions along the chromosome. An iterative weighted regression then related the trait values to the QTL genotype probabilities. The mapping steps subsequently used by Hackett et al. (2013, 2014) can be summarized as follows. The QTL genotype probabilities are estimated at each SNP from the parental genotypes and phases and the offspring dosages, using a Hidden Markov Model. QTL genotype probabilities between the SNPs are interpolated at 1 cM spacing along each chromosome, using a cubic smoothing spline. Trait values are modelled as a function of the QTL genotype, using a normal mixture model with a constant variance. As a result of computer simulations, Hackett et al. (2014) recommended fitting an additive QTL model without iterative re-weighting, and then exploring non-additive models for the genotype means estimated at the most likely position of the QTL. The SNP dosage information provides enough information to align the two parental maps so that the effects of alleles from both parents can be studied simultaneously. The analysis of the phenotypic data (lowest heritability 0.87) revealed the following QTLs (per cent of trait variance): one for maturity (51.9), one for height (35.6), two for late blight resistance of foliage (33.4, 17.8), two for late blight resistance of tubers (20.7, 16.3) and two for resistance to the white potato cyst nematode (29.8, 8.6); as well as two loci for flower colour (51.9, 7.9) and one for R-gene resistance to late blight (57.7). These all had alleles of large enough effect for them to be desirable in a new cultivar, a goal that can be achieved with the use of flanking molecular markers. The computer simulations revealed that the population size (190 offspring), and not the marker density, limited the size of QTLs that could be detected. A single QTL explaining 5 % of the trait variance in a population of 200 was detected in only 29 out of 50 simulations, and the size of the effect was on average overestimated. In contrast, with a population size of 400, such a QTL was detected in 41 out of 50 simulations, and the magnitude of the effect was estimated more accurately. It can be argued that with many alleles of small effect, association mapping is more appropriate along with genomic selection to achieve the biggest changes in phenotype possible in a particular breeding programme. With genomic selection one is more concerned with combining as many alleles with a positive effect as possible, rather than combining particular alleles.

## Examples of Polyploid Species

Having considered autotetraploid potatoes in some detail, it is now time to look at some economically important examples of triploids, allopolyploids and complex hybrids (Fig. 8.5).



**Fig. 8.5** Examples of polyploid species. (a) Triploid bananas (*Musa* spp.) on La Palma (8 February 2015). (b) Autotetraploid potatoes (*Solanum tuberosum*) near Lijiang, China (10 June 2014). (c) Allotetraploid oilseed rape (*Brassica napus*) near Edinburgh, Scotland (20 May 2012). (d) Allohexaploid wheat (*Triticum aestivum*) near Edinburgh, Scotland (14 July 2013)

## Banana and Citrus Species (Triploid)

### Banana

Bananas (*Musa* spp.) are important as the staple diet of 400 million people in Asia and Africa (Butler 2013), being grown in more than 120 tropical and subtropical countries. Today there are also important export trades, particularly from tropical America, so that bananas are grown in large, commercial plantations as well as small mixed subsistence gardens. According to Simmonds (1995b), domesticated bananas (and plantains) originated in Southeast Asia (the Malay Peninsula) and the Western Pacific, probably several millennia ago; in regions where wild species still exist in natural forests. The crop reached Africa before European contact and the Americas post Columbus.

The banana is a robust, leafy perennial herb with fleshy underground rhizomes and large leaves (the stalks form pseudostems), and massive flower clusters with separate male and female flowers. Cultivated bananas are sterile and develop seedless fruits without the need for pollination. Propagation is by division and nowadays by tissue culture. However, as the breeding is dominated by ploidy

manipulation, it will be briefly dealt with here. The interested reader is referred to the book chapter by Pillay and Tripathi (2007) for more detailed information.

The evolution of edible bananas involved human selection for parthenocarpy and sterility (Simmonds 1995b). Banana fruits are trilocular berries in which the seeds are surrounded by a mass of sweetish, starchy, parenchymatous pulp. With parthenocarpy the stimulus to pulp growth is autonomous; seeds are unnecessary. The sterility was basically genetic female sterility. In addition, gametic sterility, manifest in both ovules and pollen, was often superimposed as a result of structural chromosome changes (inversions and translocations) carried over from wild ancestors, and also accumulated during clonal propagation over long periods. Edibility first evolved in wild *M. acuminata* ( $2n = 2x = 22$ , AA genome). With high probability, several subspecies contributed and human selection favoured parthenocarpy, structural heterozygosity and seed sterility, leading to the production of edible seedless fruits, followed by male sterility consequent upon structural heterozygosity. Edible diploids are still widely but thinly distributed in Southeast Asia (New Guinea), having been largely supplanted by polyploids, primarily but not exclusively triploids ( $2n = 3x = 33$ ). Triploid plants are more vigorous and their fruits grow faster than those of diploids. There are three main groups of triploids: intraspecific (AAA) and interspecific (AAB and ABB) between *M. acuminata* and wild *M. balbisiana* ( $2n = 2x = 22$ , BB genome). All arose as a result of the production of unreduced gametes in one of the diploid parents during hybridization. Some of the triploids selected for cultivation produced somaclonal variation under continued vegetative propagation, the overall result being modern phenotypically characterized triploid subgroups of clones. Sometimes subgroups of clones have been divided into “clone sets” of similar clones.

The interested reader can find a much more detailed reconstruction of the major events in banana domestication in the perspective by Perrier et al. (2011), including geographical occurrence, timeline, and cultural associations. They integrated genetic analysis of molecular marker data from more than 400 wild and cultivated accessions of *Musa* with analysis of archaeological and linguistic data.

Banana export trades developed mainly but not exclusively in tropical America in the latter part of the nineteenth century and have depended for 100 years on remarkably few clones, all triploids (AAA), and notably Gros Michel and various mutant members of the Cavendish group. Gros Michel proved susceptible to Panama disease (banana wilt), caused by soil and water borne *Fusarium oxysporum* f.sp. *cubense*. Failure to control the disease resulted in the growing of Gros Michel being slowly wiped out between 1903 and 1960; but this also provided a stimulus to banana breeding (Simmonds 1995b). The sweeter and tastier Gros Michel was replaced by a Cavendish clone found to be resistant in 1950s. Today Cavendish accounts for 13 % of all cultivated bananas and plantains (cooking bananas) and nearly the entire export trade worth £5.3 billion. But history is repeating itself. Cavendish is susceptible to tropical race 4 (TR4) of Panama disease, which emerged in Malaysia and Indonesia in the 1990s and has now reached Mozambique and Jordan (Butler 2013). The prospect of it spreading to tropical America prompted a front page picture of bananas in The Independent Newspaper of

5 April 2014 under the heading “Bananageddon? Millions face hunger as fungus decimates global banana crop”. One possible solution is to use cisgenesis (a type of genetic transformation discussed in Chap. 17) to transfer resistance genes from cooking to dessert bananas (EFSA 2012a).

Modern banana breeding began in the 1920s but proved difficult because most of the important and popular genotypes were highly sterile and did not produce seed. As explained by Pillay and Tripathi (2007), ultimately a  $3x$ - $2x$  strategy was used in which triploid clones were crossed with productive diploids improved with resistance to pests and diseases from wild species. The method relies on a low level of seed fertility in the triploid. The crosses produced a substantial number of tetraploids (AAAA) in the progeny due to unreduced triploid gametes. Although the tetraploids had desirable characteristics, they were capable of producing seeds which is undesirable in edible bananas. Hence selected tetraploids were hybridized with improved diploids to produce secondary triploids, a  $4x$ - $2x$  strategy. In a  $2x$ - $4x$  strategy, colchicine is used to double the chromosome number of the male parent which is then crossed to diploids to produce triploids. Today banana breeding can involve parental breeding of both tetraploids and diploids.

## *Citrus Species*

*Citrus* species are diploids ( $2n = 2x = 18$ ) which were domesticated in Southeast Asia several thousand years ago and then spread throughout the world (Ollitrault and Navarro 2012). Seedlessness is one of the most important characteristics for the citrus fresh-fruit market, particularly in mandarins (*C. clementina*  $\times$  *C. tangerina*). Triploid cultivars are an efficient way to achieve this goal as they are usually both male and female sterile. Three methods of sexual triploidization have been successfully used: diploid  $\times$  natural autotetraploid crosses, autotetraploid (colchicine doubled monoembryonic diploids)  $\times$  diploid crosses, and diploid  $\times$  diploid crosses in which  $2n$  egg production occurs. The methods require embryo rescue followed by flow cytometry to recognize and select the triploids (Ollitrault and Navarro 2012). Cuenca et al. (2011) used molecular markers (SSRs) to confirm that  $2n$  egg formation in ‘Fortune’ mandarin is by second division restitution (SDR). ‘Fortune’ mandarin is widely used in triploid breeding because of its fruit qualities, late maturing period and relatively high percentage of  $2n$  eggs (6.5 %). Cuenca et al. (2011) analyzed 171 triploid hybrids from crossing ‘Fortune’ with four diploid male parents, confirmed that all of them arose from  $2n$  megasporangia in ‘Fortune’, inferred the genotypes of the  $2n$  gametes, and ruled out first division restitution (FDR) from the pattern and rates (<50 % for several markers) of restitution heterozygosity.

## Brassicas (Diploids and Allotetraploids) and Raphanobrassica

The economically most important Brassicas comprise three diploid and three tetraploid species. The diploids are *B. rapa* (turnip, turnip rape, Chinese cabbage and Chinese mustard or Pak-choi), *B. nigra* (black mustard) and *B. oleracea* (kale, cauliflower, cabbage, Brussels sprouts, kohlrabi and broccoli), and the tetraploids are *B. juncea* (Indian or brown mustard), *B. napus* (oilseed rape, forage rape and swedes) and *B. carinata* (Ethiopian mustard). Information on their histories as crops can be found in the second edition of *Evolution of Crop Plants* by Smartt and Simmonds (1995).

The foundations to our understanding of the genetic relationships between the six Brassicas were laid by the Korean-Japanese botanist Nagaharu U, who made and examined chromosome pairing in synthetic hybrids between the diploid and tetraploid species (Nagaharu 1935). His resulting theory is known as the “Triangle of U”. He correctly proposed that the three tetraploid species *B. juncea* (AABB,  $2n = 36$ ), *B. napus* (AACC,  $2n = 38$ ) and *B. carinata* (BBCC,  $2n = 34$ ) are the allotetraploids of the diploid species *B. rapa* (AA,  $2n = 2x = 20$ ), *B. nigra* (BB,  $2n = 2x = 16$ ) and *B. oleracea* (CC,  $2n = 2x = 18$ ). The diploid species are outbreeders with a sporophytic self-incompatibility system. The allotetraploids arose by natural hybridization and chromosome doubling, should also be self-incompatible, but are usually self-compatible and tolerant of inbreeding. Today there is interest in widening their genetic base by resynthesis from their progenitors, facilitated by embryo rescue and other techniques.

## Arabidopsis–Brassica Genome Evolution

Starting in the 1990s, a combination of research on the model plant *Arabidopsis thaliana* and *Brassica* genetics has revealed additional complexity in the evolution of the diploid plant genomes. There is now good evidence that the three basic genomes contain three partially rearranged copies of a much older genome that was very similar to that of *A. thaliana*, which itself is almost certainly an ancient tetraploid. There is in fact evidence for additional large-scale gene duplications in its evolutionary past (Simillion et al. 2002). Murphy (2007) summarized possible events as follows. Over 40 million years ago there was a small plant of the mustard family with a tiny genome of about 60 Mb. About 38 million years ago this plant either spontaneously doubled its chromosomes or hybridized with a close relative to create a new tetraploid of size 120 Mb. Modern *A. thaliana* is the direct descendant of this plant but behaves like a conventional diploid. About 20 million years ago some of the *Arabidopsis*-like tetraploid plants formed new polyploid hybrids that initially contained two and then three ‘tetraploid’ genomes of total size

250–400 Mb. Some 5 million years ago one of the new hybrids with three ‘tetraploid’ genomes became a successful new species of size 500–600 Mb and was the progenitor of today’s ‘diploid’ brassicas through functional diploidization. About 4 million years ago this progenitor diverged into several closely related species including *B. rapa*, *B. nigra* and *B. oleracea*. Finally, as recently as 2000 years ago, today’s ‘tetraploid’ brassica crops arose by interspecific hybridization. Oilseed rape (*B. napus*) has a genome size of 1200 Mb. The work of Ziolkowski et al. (2006) and others has revealed that since the tetraploidization of the *A. thaliana* ancestor, chromosomal rearrangements have modified its genome evolution. Ziolkowski et al. (2006) used FISH analysis of chromosomal breakpoints to demonstrate that the majority of rearrangements occurred before the *Arabidopsis*–*Brassica* split and that the segmental architecture of the *A. thaliana* genome is predominantly conserved in *Brassica*. They also concluded that the chromosomal rearrangements were an important part of the diploidization process and that they occurred shortly after the polyploidization event. Beilsteina et al. (2010) used previously overlooked fossil evidence and phylogenetic analysis to date the split between *Arabidopsis* and the *Brassica* complex to 43 million years ago; pushing the duplication event that led to their progenitor back to around 54 million years ago. Beilsteina et al. (2010) confirmed that the whole genome triplication occurred about 20 (22.5) million years ago. Much more detailed analysis of *Brassica* genome evolution should be possible now that draft sequences of the *B. rapa* (*Nature Genetics*, 28 August 2011) and *B. oleracea* (*BMC Genomics*, 30 September 2013) genomes are available. Furthermore, analysis of the grapevine (*Vitis vinifera*) genome (*Nature*, 26 August 2007) provided evidence of palaeo-hexaploidy (i.e. three ancestral genomes) that was present in the common ancestor to *Arabidopsis* and grapevine, and therefore a trait common to all Eurosids. The triplicated arrangement is apparent after the separation of the monocotyledons and dicotyledons and before the spread of the Eurosid clade. Future genome sequencing projects for other clades of dicotyledons, such as Solanaceae or basal eudicots, will help in situating the triplication event more precisely, and eventually in establishing its precise nature (hexaploidization or genome duplications at distant times). No doubt the results will be fascinating, but it is too early to assess their impact on crop improvement.

## ***Raphanobrassica***

Before leaving Brassicas, Raphanobrassica deserves a mention. Raphanobrassica (RRCC,  $2n = 36$ ) is the intergeneric hybrid between radish (*Raphanus sativus* RR,  $2n = 2x = 18$ ) and kale (*Brassica oleracea* CC,  $2n = 2x = 18$ ). As far as is known, it does not occur naturally. Nevertheless, it has been successfully synthesized both by colchicine treatment of seedlings of diploid hybrids and by crossing autotetraploid forms of the parent species, but only with *R. sativus* as female parent (McNaughton 1973). Doubling of diploid hybrids results in completely

homozygous amphidiploids whereas crossing at the tetraploid level using heterozygous parents produces genetically variable offspring. Ellerström and Zagorcheva (1977) synthesized Raphanobrassica from autotetraploid fodder radish and marrowstem kale. Through repeated selections over six generations they were able to increase seed setting from 0.1 to 2.3 seeds per pollinated flower, which is equivalent to a seed fertility of 13 %. They concluded that prospects were good for Raphanobrassica as a new forage crop provided an economic seed production could be achieved, which would require a fertility of around 40 %. A decline in the use of forage brassicas meant that there was insufficient breeding effort to achieve this goal, but it remains an interesting prospect.

### ***Festolium* (Allotetraploid)**

*Festolium* refers to natural or synthetic intergeneric hybrids between obligate outbreeding species of the *Festuca* (fescue) and *Lolium* (ryegrass) genera (Ghesquière et al. 2010). These species are frequently considered ideal components of grassland for agricultural and amenity purposes. Ryegrasses provide high yields of nutritious forage but lack the resilience against abiotic stresses found in the more robust and stress-adapted fescues; hence the interest in their hybridization. The following brief account is based on the review by Ghesquière et al. (2010). The first synthetic *Festolium* hybridization was reported by Jenkin (1933) and involved perennial ryegrass (*L. perenne*) and meadow fescue (*F. pratensis*). However, the commercial breeding of *Festolium* cultivars developed much later. In the European Union, *Festolium* was initially defined as the hybrid between Italian ryegrass (*L. multiflorum*) and meadow fescue (*F. pratensis*), but was later extended to include all intergeneric hybrids resulting from crossing a species of *Festuca* with one of *Lolium*, regardless of chromosome number and whether hybrids were intentionally backcrossed into their parental species.

If we stick with the initial narrow definition of *Festolium*, hybrids were produced by taking diploid *L. multiflorum* ( $2n = 2x = 14$ ) and diploid *F. pratensis* ( $2n = 2x = 14$ ), treating them with colchicine to produce the autotetraploid forms ( $2n = 4x = 28$ ), and making the cross at the tetraploid level. The intact genomes of the two species were therefore combined, and as preferential homologous chromosome pairing occurred, chromosome composition was assumed to be stabilized across generations; in other words, amphidiploidy and disomic inheritance were established. However, this may not be strictly true. A significant and progressive loss of *Festuca* chromosomes was demonstrated in an autotetraploid *F. pratensis* × autotetraploid *L. perenne* cross; from 14 *Lolium* + 14 *Festuca* chromosomes in the  $F_1$  hybrid to an  $F_6$  plant with 19 *Lolium* + 9 *Festuca* chromosomes including recombinant chromosomes (Ghesquière et al. 2010). These features were revealed by genomic in situ hybridization (GISH) in which *Lolium* and *Festuca*

chromosomes were stained different colours. Crossing can also be done at the diploid level followed by *in vitro* rescue of hybrid embryos 10–16 days after pollination. The *Lolium* species is commonly used as the female parent. The resulting hybrids are then treated with colchicine to double the chromosome number and restore fertility. The hybrids can be used as parents in further breeding leading to cultivars. Other *Festuca* species are polyploid with homoeologous genome sets. Despite their incomplete homology, homoeologous *Festuca* chromosomes do pair preferentially in *Festulolium* because of the greater chromosome pairing affinity of the disomic sets of homologous *Lolium* chromosomes. Thus instances of intergeneric chromosome pairing are restricted, but strict disomic inheritance may not be fully achieved.

Twenty-three tetraploid *Festulolium* cultivars are mentioned by Ghesquière et al. (2010), of which 16 were from autotetraploid *L. multiflorum* × autotetraploid *F. pratensis* crosses, or reciprocal crosses. The benefit of tetraploid *Festulolium* cultivars lies in their potential to achieve an optimum balance in productivity, quality, and stress tolerance that is not achievable by conventional intraspecific breeding. However, seed yield in multiplication fields relative to the pure parent species is still to be confirmed with accurate assessments. It is therefore too early to predict the likely future impact of *Festulolium*.

## Cereal Genome Evolution

Before moving on to wheat it is worth mentioning a few points about cereal genome evolution. They will also be relevant to the final section on inbred lines of maize. Murphy (2007) summarized the timescale of the evolution of cereals, based on genome sequence comparisons and the molecular clock, as follows. The monocots diverged from the dicots about 200 million years ago (MYA). The first grass species date from about 100 MYA and rices appeared about 40 MYA and evolved in South and East Asia. The ancestors of maize and sorghum split 15 MYA and evolved in Mesoamerica and North Africa, respectively. Wheat, barley and rye diverged from a common ancestor about 10 MYA in the Near East. Wheat and rye diverged 7 MYA and *Aegilops* and *Triticum* 3 MYA. In contrast, domestication began about 13,000 years ago. Cereal genome evolution was in fact a complicated process. Murat et al. (2014) have proposed a scheme that is compatible with the latest genome sequence data. It starts with an ancestral grass karyotype of seven protochromosomes containing 6045 ordered protogenes that underwent whole genome duplication (about 50–70 million years ago) followed by ancestral chromosome fusion. Hence almost all modern diploid grass species are paleopolyploids. We will end this chapter by briefly exploring what happened in maize; but first we will look at how hexaploid wheat arose.

## Wheat (Allohexaploid)

The evolutionary history of allotetraploid and allohexaploid wheat was reviewed by Levy and Feldman (2004), based on molecular data such as the gene sequence comparisons done by Huang et al. (2002). Wild diploid wheats from the *Triticum-Aegilops* group diverged from a common diploid progenitor about 4.5 million years ago. Then about 0.5 million years ago, intergeneric hybridization between diploid *T. urartu* (genome AA,  $2n = 2x = 14$ ) as male and an unknown diploid species similar to *Ae. speltoides* (BB,  $2n = 2x = 14$ ) as female, followed by chromosome doubling, gave rise to the wild allotetraploid wheat, *Triticum turgidum* ssp. *dicoccoides* (AABB,  $2n = 4x = 28$ ) (Levy and Feldman 2004). This species was the direct progenitor of cultivated emmer (ssp. *dicoccoides*) and durum (ssp. *durum*) wheats, domestication taking place about 10,500 years ago. Then about 9500 years ago, a second round of intergeneric hybridization and chromosome doubling took place, this time between the domesticated allotetraploid wheat as female and the wild diploid species *Ae. tauschii* (DD,  $2n = 2x = 14$ ) as male. The result was allohexaploid bread wheat *T. aestivum* ssp. *aestivum* (AABBDD,  $2n = 6x = 42$ ). Levy and Feldman (2004) also describe the complex genetic and epigenetic alterations that occurred in wheat on polyploidization, the outcome of which was a stable polyploid which displays disomic inheritance. The alterations comprise non-random elimination of coding and non-coding DNA sequences and changes in gene expression as a result of epigenetic changes (e.g. DNA methylation of coding and non-coding DNA) and activation of retroelements. The details need not concern us here, but it is fair to say that the wheat genome really is reprogrammed. Despite this complexity, it is still true to say that hexaploid wheat behaves as a diploid at meiosis due to the *Ph1* (pairing homoeologous) gene on the long arm of chromosome 5B, which was thought to prohibit pairing between homoeologous chromosomes (Riley and Chapman 1958). Recently Martín et al. (2014) have shown that *Ph1* stabilises polyploidy in wheat by both promoting homologue pairing and preventing MLH1 (a DNA mismatch repair protein) sites from becoming crossovers on paired homoeologues during meiosis. They were able to do this by studying meiosis in wheat-rye hybrids which have 28 homoeologues but no homologues. Since the absence or mutation of *Ph1* allows homoeologous pairing and recombination, manipulation of this gene has been a major approach for transferring desirable genes from related species to common wheat. For example, a large number of translocation lines containing alien genes for resistance to abiotic and biotic stresses have been developed using allele *ph1b* (Mergoum et al. 2009a). One further point worth making is that successful intergeneric hybridizations took place on many occasions, thus resulting in genetic variation in the newly formed species (Murphy 2007).

## Triticale: A New Crop

The history of what became a new crop, triticale, can be found in the book chapter by Mergoum et al. (2009b) in Volume 3 of the Handbook of Plant Breeding, as well as in an earlier review by Oettler (2005). Here a brief summary will suffice. The first fertile triticale was produced by Rimpau in 1888 from crosses between *Triticum aestivum* (common wheat) and rye (*Secale cereale*), followed by spontaneous chromosome doubling. Isolated research on triticale occurred in Europe over the next 50 years but did not result in a commercial crop. It was breeding programmes in the 1950s and 1960s that produced the first cultivars, from Hungary in 1968 and from Spain and Canada in 1969, and led to the development of triticale as a cereal crop. The efficiency of production of the intergeneric hybrids was improved by using *in vitro* culture of the hybrid embryos to increase their survival rate, followed by the use of colchicine on the resulting plants to ‘double their chromosomes’ before flowering, and hence ensure their fertility. Today triticale is a well-established crop internationally with 13.5 million tons of grain harvested in 28 countries across the world in 2005, the leading producers being Germany, France, Poland, Australia, China and Belarus; but figures from North America were missing. Triticale is used as a food grain for humans (bread, pasta and breakfast cereal) and as a feed grain for animals, both monogastric and ruminant. It is grown as a forage crop for livestock grazing or conservation as whole-plant silage and hay. It is also attracting attention as a potential energy crop.

Two basic types of primary triticale have been produced, hexaploid and octoploid. The seed parent of the former is durum wheat (*Triticum durum*,  $2n=28=AABB$ ) and of the latter is common wheat (*Triticum aestivum*,  $2n=42=AABBDD$ ). Cultivated diploid rye (*Secale cereale*,  $2n=14=RR$ ) is the pollen parent of both. Hence the genome constitutions of hexaploid and octoploid triticales are ( $2n=42=AABBRR$ ) and ( $2n=56=AABBDDRR$ ), respectively. Both allopolyploids are amphidiploids (diploid inheritance) which self-pollinate like their wheat parent. Although octoploid triticales were the first to be produced and extensively studied, it was the hexaploids which had greatest commercial success due to their superior vigour and greater reproductive stability. The vast majority of today’s triticales are descendants of these primary ones. Further chromosome engineering has resulted in some chromosomes from the R genome being replaced by some from the D genome; for example following  $AABBDD \times AABBRR$  crosses. Although triticale is an inbreeding crop, a yield advantage of about 20 % has been found in hexaploid hybrids produced using a chemical hybridizing agent to induce male sterility in one parent. Achievements in triticale breeding have been reviewed by Mergoum et al. (2009b) and include improvements in yield and quality, adaptation to abiotic stresses and resistance to biotic stresses.

## Complex Polyploids: Sugarcane

Sugarcane, *Saccharum officinarum* ( $2n = 80$ ), or “noble cane” due to the sweetness of its stalk juice, is a domesticated tropical perennial grass species which is propagated vegetatively from stem cuttings. Modern cultivars have a large range of chromosome numbers ( $2n = 100\text{--}125$ ) and a genome sequence of about 10 Gb (Scortecci et al. 2012). They have a complex evolutionary history involving polyploidization, interspecific hybridization, and partial loss of chromosomes (aneuploidization). However, the basic sugarcane haplotype ( $x = 10$ ; 930 Mb) is relatively small and syntenic to model grasses such as sorghum (730 Mb); sugarcane and sorghum lineages having diverged 5 million years ago (Paterson et al. 2004). Sugarcane is one of six commonly recognized species of *Saccharum*, although the species *S. edule*, cultivated for its edible inflorescence in New Guinea and the Fiji Islands, need not concern us. The six species are largely cross-pollinated by wind and suffer inbreeding depression. They are polyploids with very high chromosome numbers and considerable amounts of aneuploidy. Polysomic inheritance predominates. The six species can be successfully crossed, and they have limited interfertility with other genera, including *Sorghum*. A summary of the complex taxonomy, characteristics and evolution of sugarcane species can be found in the review by Roach (1995) and in Hancock (2012).

According to Roach (1995), sugarcane was probably first domesticated from the wild species *S. robustum* ( $x = 10$ ;  $2n = 80$ ; but variation from 63 to 205) in or near New Guinea a few thousand years ago. Human selection for chewing plants with sweet juice and low fibre presumably produced *S. officinarum* ( $x = 10$ ;  $2n = 80$ ; 70–140), with its centre of diversity in this area. Cultivation moved northwest into Asia, where hybridization with wild *S. spontaneum* (an autopolyploid;  $x = 8$ ;  $2n = 64$ ; 32–128) produced cultivated *S. barbieri* (Indian cane;  $2n = 81\text{--}124$ ) and *S. sinense* (Chinese cane;  $2n = 115\text{--}120$ ). These hybrid “thin” canes were less sweet and less robust than the noble canes, but were well adapted to the seasonal monsoon climate of northern India and southern China. They provided the basis for local syrup and crude sugar production some 3000 years ago. Further dispersal of these took place to southern Europe and northern Africa, as well as *S. officinarum* eastwards across the Pacific. From the sixteenth century, sugar production for world trade changed progressively from cottage industries based on *S. sinense* and *S. barbieri* to plantation and factory industries based on noble canes. Sugarcane was introduced to America by Columbus on his second journey (1493). The Portuguese introduced sugarcane to Brazil during the sixteenth century, probably with hybrids between *S. officinarum* and *S. barbieri* originated from India and Iran; and today Brazil is the largest producer in the world. The crop was propagated by stem cuttings for the next 250 years, and this vegetative reproduction allowed the accumulation of numerical and structural chromosome changes (Cuadrado et al. 2004).

Continuing from Roach (1995), recognition in 1888 of the ability of sugarcane to produce true seed marked the beginning of modern breeding as an alternative to

cultivar collection and selection. Early breeding initially relied on intraspecific crossing (open-pollination) of noble canes and resulted in better sugar yield. However, from 1893 the emphasis switched to hybridizations with the wild species *S. spontaneum* and selection for disease resistance as well as sucrose yield. Success came in 1921 with the breeding of the “wonder cane” POJ2878, usually referred to as a “nobilized cane”. It is present in the pedigrees of most modern cultivars. These are essentially derivatives of *S. spontaneum* backcrossed to noble types. When the noble type is the female parent in the initial cross, there is evidence of  $2n$  female gametes combining with  $n$  male gametes to form the hybrid. Progeny of back-crosses to noble canes have, on average, fewer chromosomes than expected, and commercial hybrids have chromosome numbers in the range  $2n = 100–125$ . From the 1930s “nobilized canes” were intercrossed to produce “hybrid” cultivars which in turn, from the 1950s, were intercrossed to give modern cultivars (Ming et al. 2006). Thus modern sugarcane cultivars are complex hybrids synthesized principally from noble canes and *S. spontaneum*, with minor contributions from *S. sinense* and *S. barberi*, and to a lesser extent, *S. robustum*. New cultivars are derived from extensive recombination and selection among materials of this kind.

Modern cytogenetics has revealed interesting details about the composition of complex hybrid cultivars. The long-debated size of sugarcane’s basic chromosome set was established by the physical mapping of ribosomal RNA genes. *S. officinarum*, *S. robustum* and *S. spontaneum* have basic chromosome numbers of  $x = 10$ ,  $x = 10$  and  $x = 8$ , respectively (D’Hont et al. 1998). As a consequence, two distinct chromosome organizations coexist in current hybrid cultivars. Genomic *in situ* hybridization (GISH) has been used to identify parental chromosomes in interspecific hybrids through incorporation of different coloured dyes, and for testing for the exchange of material between genomes. Among the chromosomes of cultivar ‘R570’ ( $2n = 107–115$ ), about 80 % were identified as originating from *S. officinarum*, 10 % as originating from *S. spontaneum* and another 10 % as recombinant chromosomes, demonstrating that exchanges have occurred between chromosomes derived from *S. officinarum* and *S. spontaneum* (D’Hont et al. 1996). In further work, Cuadrado et al. (2004) compared three sugarcane cultivars, ‘My5514’, ‘B42231’ and ‘C236-51’, with a typical *S. officinarum*, ‘Cristalina’ which produces the highest sugar yield but is highly susceptible to sugar cane mosaic virus (SCMV). The commercially successful cultivar ‘My5514’ has good resistance to SCMV and fungal (*Ustilago scitaminea*) infections. All hybrid cultivars were found to be aneuploid, affecting both parental genomes (having chromosomes in addition to full genomes), with chromosome numbers from  $2n = 102–106$  in ‘My5514’ and up to  $2n = 113–117$  in ‘C236-51’. Among the chromosomes of the three hybrids, about 16 % were identified as having been contributed by *S. spontaneum*, and less than 5 % as recombinant and/or translocated chromosomes between *S. officinarum* and *S. spontaneum*. These chromosomes could have arisen from recombination events in the few meiotic opportunities that occurred in the interspecific hybrids and their early generation descendants. Structural interchanges might also be explained by spontaneous translocations accumulated during vegetative propagation. However, the origin of interspecific chromosome interchanges

cannot be deduced from karyological observations, and chromosome pairing behaviour has not been definitively clarified in sugarcane. Analysis of rDNA (ribosomal DNA) provided additional information, but need not concern us here. More information can be found in the reviews by Ming et al. (2006) and Scortecci et al. (2012).

## Inbred Lines of Maize and the Pan-Genome Concept

Next-Generation Sequencing (NGS) methods have allowed the genomes of individuals within a species to be compared. Maize has attracted a lot of attention because the same is true of its inbred lines used to produce commercial F<sub>1</sub> hybrid cultivars. Naturally comparisons have been made with inbred line B73 as this was the first to be sequenced (published in *Science* on 20 November 2009), and comprised 2.3-Gb (gigabase pairs, i.e. 2.3 billion) with a predicted 32,540 protein-coding genes. Sequencing B73 revealed that 85 % of its genome is composed of hundreds of families of transposable elements (TEs), dispersed nonuniformly across the genome. Transposable elements can move around the genome and are of interest to breeders because they cause mutations. Here, however, I want to consider what the research done since 2009 tells us about genome evolution in maize, its high levels of both phenotypic and genetic diversity, and the possible consequences for crop improvement. The ‘Science’ paper made the following points about knowledge in 2009. The maize genome has undergone several rounds of genome duplication, including that of a paleopolyploid ancestor about 70 million years ago (MYA), and an additional whole-genome duplication event about 12 to 5 MYA, which distinguishes maize from its close relative, *Sorghum bicolor* whose genome sequence was also published in 2009 (in *Nature* on 29 January 2009). The sorghum genome ( $2n = 2x = 20$ ) comprises 730-Mb (megabase pairs, i.e. 730 million) with a gene order and density similar to rice. Retrotransposon accumulation in recombinationally recalcitrant heterochromatin explains the 75 % larger genome size of sorghum compared with rice. Although gene and repetitive DNA distributions have been preserved since the palaeopolyploidization about 70 million years ago, most duplicated gene sets lost one member before the sorghum–rice divergence. In contrast, after the duplication event ( $2n = 4x = 40$ ) about 12 to 5 MYA, the return to a genetically diploid state ( $2n = 2x = 20$ ) was associated with numerous chromosomal breakages and fusions, as shown by alignment to the genomes of sorghum and the more distantly related rice. As sorghum experienced relatively few interchromosomal rearrangements since its lineage split with rice, its chromosomal configuration closely resembles the ancestral state of maize’s two subgenomes. Cosynteny of maize genes to common reference genes in rice or sorghum, define maize’s duplicate regions. Although syntenic blocks cover 1832 Mb (89 % of the genome), individual gene losses were common and resulted in retention of only about 8110 genes as duplicate homoeologues. In other words, there was a strong bias for gene loss (fractionation)

between sister regions. Nevertheless, over the last 3 million years, the maize genome expanded to its current size via a proliferation of long terminal repeat retrotransposons (LTR retrotransposons). Some of these changes are reflected in differences between the inbred lines of maize which have been used to produce F<sub>1</sub> hybrid cultivars. Hence they are of interest to maize breeders.

### ***Copy Number and Present/Absent Variants***

Resequencing and array-based comparative genomic hybridization between B73 and inbred line Mo17 revealed extensive structural variation, including hundreds of copy number variants (CNVs) and thousands of present/absent variants (PAVs). Many of the PAVs contain intact, expressed single-copy genes that are present in one inbred genome but absent from the other. Springer et al. (2009) used whole-genome, array-based, comparative genomic hybridization (CGH) to determine the extent of structural variation between B73 and Mo17. They identified at least 2056 unique PAVs, of which 1270 could be genes (4 % of total maize genes), and more than 400 putative CNVs which were likely to include functional genes. In addition there were large regions (>1 Mb) that had little or no variation. The authors estimated that over 20 Mb of DNA present in B73 was absent from Mo17, some of which could possibly be uncharacterized transposable elements. In conclusion, the authors thought that the high level of PAVs reflected ancient haplotype variation rather than recent novel inbred-specific events. A number of researchers have suggested that these haplotype-specific sequences may contribute to heterosis (hybrid vigour) in maize and the substantial degree of phenotypic variation among maize inbreds. In other words, they have important implications for maize improvement.

### ***Pan-Genome Concept***

Earlier, Morgante et al. (2007) had found that transposable elements were largely responsible for extensive variation in both intergenic and local genic content among individuals within a species. They concluded that a single genome sequence might not reflect the entire genomic complement of a species and introduced the concept of the plant pan-genome. This includes core genomic features that are common to all individuals in a species and a dispensable genome composed of partially shared and/or non-shared DNA sequence elements. Comparing four orthologous genomic regions of B73 and Mo17, they estimated a pan-genome comprising a shared core genome of 1.67 Gb (50 %) and a dispensable genome of the same total size that is equally distributed among the two lines. They thought that the core genome comprised both single-copy sequences (including most if not all genes) and transposable elements that are found among all individuals in a certain genomic location.

In contrast, the dispensable genome was mostly made up of transposable elements of different types that were present in multiple copies in each individual but which could be found in a specific location only in some of them. The differences between B73 and Mo17 in intergenic regions were accounted for by the transposition of several different families of retroelements. These retroelements were present in specific inbred lines, and had been inserted significantly more recently than the shared retroelements; the majority within the past 400,000 years.

Lai et al. (2010) extended resequencing of maize inbred lines to a comparison of Zheng58, 5003, 478, 178, Chang7-2 and Mo17 with B73. Lines Zheng58, Chang7-2, 178 and Mo17 are all members of a popular heterotic group used in China (Mo17 is also a member of an important heterotic group used in the USA). Zheng58 and Chang7-2 are the parents of the commercial hybrid (ZD958) that is currently the most widely grown one in China. Inbred line 178 is the female parent of another hybrid (ND108) that is also widely grown in China. Inbred line 478 is a parent and inbred line 5003 is a grandparent of Zheng58. The authors identified 1,272,134 SNPs in non-repeat regions, with 468,966 in the 32,540 ‘high-confidence’ maize genes (there is still uncertainty over exact number of genes) and 130,053 SNPs in coding regions. They also identified 30,178 indels ranging from 1 to 6 bp in length, of which 571 were in coding regions. They found 101 genomic intervals (average length 2.4 Mb) scattered throughout the genome that had low sequence diversity and which contained genes known to have been under selection during maize improvement. Compared with B73, they found 104 regions (5 kb or longer) which could be regarded as putative missing genes in Mo17. In total they found 296 ‘high-confidence’ genes (i.e. intact expressed genes) in B73 that were missing from at least one of the other six inbred lines. Most of the PAV events seemed to involve only a single gene, but some included deletions of two to four adjacent genes and one large deletion (2 Mb) on chromosome 6 of Mo17 had at least 18 genes deleted.

### ***Maize Subgenomes***

Schnable et al. (2011) used comparative analysis of the maize and sorghum genomes to examine the differentiation of duplicated genomic regions following the maize tetraploidy (either autotetraploidy or allotetraploidy) event 12 to 5 MYA. Today maize-1 subgenome is much larger than maize-2 (1.26 versus 0.75 Gb) and it would appear that biased fractionation acted on all genomic DNA and not simply on coding sequences; but this was not proved. The authors found a pattern of overexpression of genes from the genome that had experienced less gene loss. Hence there appears to have been selection against loss of the gene responsible for the majority of the total expression for a duplicate gene pair. The authors also suggest that the gene contents of maize-1 and maize-2 genomes were already significantly different at the time of this most recent transposon insertion and expansion of the past few million years. The transposons inserted into maize-1

and maize-2 in approximate proportion to the gene content of these regions, thus magnifying the differences between the subgenomes. Although the tetraploidy of maize is ancient, biased gene loss and expression have continued to the present day and explain, at least in part, the remarkable genetic diversity found among modern maize cultivars. Using only maize genes with retained syntenic orthologues in both sorghum and rice, the authors constructed two lists of high-confidence genes, the list of retained homoeologues from the maize duplication and the list of genes for which it was possible to say with high confidence that the duplicated copy was lost from the genome (singleton genes). The lists were further subdivided into maize1-specific and maize2-specific lists of genes.

### ***Maize Pan-Transcriptome***

Hirsch et al. (2014) used transcriptome sequencing of seedling RNA from 503 maize inbred lines to characterize the maize pan-transcriptome as a proxy for the maize pan-genome. They identified 8681 representative transcript assemblies (RTAs), defined as the longest transcript within a locus. Of these, 16.4 % were expressed in all lines and 82.7 % expressed in subsets of the lines. They worked out that 350 of the lines were adequate to represent all of the genes. The authors consider the RTAs identified in their study to represent some of the genes thought to be absent in one or both of the maize subgenomes. They thus confirmed that a substantial portion of variation (genes) in a species may lie outside a single reference genome for a species. References in their paper can be found to further results on maize and from other plant species.

## **Conclusions for Plant Breeders**

The clear message for plant breeders is that the evolutionary histories of their crops, over millions of years and not just since domestication, tell the stories of how the variation seen today in their genomes came into existence. It is this variation that breeders need to continue to use to produce tomorrow's cultivars. In Part III we will explore the methods they have used to date before considering continuing and new options for the future in Part IV.

**Part III**

**Landraces to High Yielding Cultivars**

# **Chapter 9**

## **Genetic Structure of Landraces**

### **Introduction**

Domestication, geographical dispersal to new environments, and selection by farmers and growers, resulted in numerous landraces and strains of cultivated plant species. Brief accounts of these processes for over 200 crops can be found in the second edition of *Evolution of Crop Plants* by Smartt and Simmonds (1995), and further information for many crops in the Handbook of Plant Breeding (Prohens et al. 2008). Furthermore, genetic analyses of extant landraces are allowing inferences to be made about the evolutionary histories of crops. For example, Jones et al. (2011) analyzed 651 barley landraces from across Europe with 24 microsatellites and found that they could be grouped by STRUCTURE analyses into 8–11 populations with different geographical distributions. Population structure was associated to varying extents with phenotype, climate and spread through Europe. For now, however, we can accept that the genetic differences which arose between and within landraces and strains were the result of the processes studied by population and evolutionary geneticists (see for example, *Principles of Plant Breeding, Second Edition* by Allard (1999), *Elements of Evolutionary Genetics* by Charlesworth and Charlesworth (2010) and *Genetics of Populations, Fourth Edition* by Hedrick (2011)). We need to concentrate on the genetic structure of individual landraces and strains in order to understand how, during the last 200 years, different types of cultivars have been bred from this germplasm. We shall find that the type was determined by the mode of reproduction (sexual or asexual) and mating system (extent of self- versus cross-pollination) of the cultivated plant species. We shall assume that the landrace is a large, isolated population with no gene (pollen) flow from a neighbouring landrace and that any new mutations will be rare so that they can be ignored in the short term. We will consider what happens over many generations in the absence of both natural and

farmer selection, but will demonstrate that the effects of past selection cannot be ignored. Furthermore, as most plants are hermaphrodite, we will assume a mixture of self- and cross-pollination, the latter occurring at random with respect to genotype (Bennett and Binet 1956). Modern studies of *in situ*, on-farm, conservation of biodiversity in crop plants indicate that our assumptions are a simplification. Crop populations should probably be considered metapopulations, which are grown by communities of farmers, with subpopulations (landraces) managed by individual farmers in spatially discrete fields. There is likely to be seed exchange between farmers, and furthermore, the farmers are likely to grow mixtures of what they regard as recognizable varieties. Genetic drift (random changes) will be operating when farmers select small numbers of seed for planting the next generation. An example is the sorghum populations managed by Duupa farmers in northern Cameroon (Alvarez et al. 2005). The situation is even more complicated with populations of most clonal crops because they consist of two components, established clones and sexually produced volunteers (McKey et al. 2010). Nevertheless, our simplified models of landraces should lead to some useful conclusions.

Interestingly, in their analysis of 203 crops, Meyer et al. (2012) found considerable variation in reproductive strategies for a number of crops: 25 had both self-fertilizing and outcrossing strategies, 32 both self-fertilizing and vegetative propagation strategies, and 66 both outcrossing and vegetative propagation strategies. Overall, 95 crops could be classified as self-fertilizing, 102 as outcrossing and 115 as having vegetative reproduction. A number of crops reviewed (27 %) exhibited a change in reproductive strategy between their wild and domesticated forms, either from outcrossing to self-fertilizing or from sexual reproduction to vegetative propagation. This change was considered an important feature of crop domestication because it was a mechanism for establishing reproductive isolation, allowing farmers to maintain desired phenotypes. However, most self-fertilizing crops (over 80 %) were already self-fertile in their wild state, presumably making them favourable candidates for domestication.

## Mixtures of Inbred Lines, Hybrids and Clones

Before considering the genetic structure of individual landraces and strains in detail, it is useful to think of landraces as mixtures of inbred lines, or hybrids, or clones, depending on their mode of reproduction and mating system (Fig. 9.1).



**Fig. 9.1** Mode of reproduction and mating system. (a) Wheat: self-pollinated. (b) Maize: cross-pollinated (separate male and female flowers). (c) Oilseed rape and Faba bean: insect-pollinated, mixture of self- and cross-pollination. (d) Potato: sexual reproduction and clonal propagation

### ***Sexual Reproduction with Predominantly Self-Pollination***

Crops such as wheat, rice and barley normally reproduce by self-pollination and the level of outcrossing is usually less than 5 %. For example, Abdel-Ghani et al. (2004) reported rates of 0–1.8 %, with a mean of 0.34 %, in 12 barley (*Hordeum vulgare*)

landraces and 13 wild barley populations in Jordan. In tomato (*Solanum lycopersicum*) the outcrossing rate in three outdoor cultivars in 2 years in Central Germany ranged from 0 to 4 % and could be related to the extent of stigma protrusion from the anther cone (Horneburg and Myers 2012). In other words, self-pollination is enhanced through the enclosure of the stigma and style by the anther cone which protects them from other sources of pollen. Some crops which we think of as predominantly self-pollinators do sometimes have higher outcrossing rates; for example, the common bean (*Phaseolus vulgaris*). Ibarra-Perez et al. (1997) found average outcrossing rates of 6.9 % for six lines of common bean representing Mexican germplasm. However, the range for individual maternal lines (progenies) was 0.0–78.0 %, and showed the importance of environmental factors through interactions of lines with location  $\times$  planting date and location  $\times$  year.

If 100 % self-pollination occurred, the landrace population would consist of a mixture of true breeding inbred lines. Even allowing for 5 % cross-pollination, this is still true, but the 95 % of the population that reproduced by self-pollination in any generation will be a mixture of inbred lines at different levels of inbreeding. This is easily seen by considering what happens over many generations, say  $n+1$ . In the first generation, 95 % of the population will have arisen by self-pollination and 5 % by cross-pollination. After a further  $n$  generations, that is  $n+1$  generations in total,  $(0.95) \times (0.95)^n$  will be the proportion of the population that has been self-pollinated for  $n+1$  generations, and  $(0.05) \times (0.95)^n$  will be the proportion that has been self-pollinated for  $n$  generations. Therefore  $(0.95 + 0.05) \times (0.95)^n = (0.95)^{n+1}$  will be the proportion of the population that has been self-pollinated for at least  $n$  generations. The division each generation into 95 % self-pollination and 5 % cross-pollination does not affect this simple result as shown in Table 9.1 for five generations of self-pollination. Extension of the table to many generations would show that half of the population will comprise inbred lines that have been self-pollinated for 13 or more generations.

### ***Sexual Reproduction with Predominantly Cross-Pollination***

Maize normally reproduces by cross-pollination and the level of self-pollination is usually less than 5 %. Although there is no physiological mechanism to prevent self-pollination, maize is a wind-pollinated monoecious species with staminate male flowers at the top of the plant (tassels) and pistillate female flowers half way down the stalk (silks are long styles). Furthermore, maize is protandrous, with the pollen shed before silk emergence. Most cross-pollination is from neighbouring plants, but some is long distance. If there was 100 % cross-pollination, each plant would be the offspring of two parents and could be regarded as a hybrid of these two parents. Hence a landrace population can be viewed as a mixture of hybrids. Even

**Table 9.1** Levels of inbreeding in population after five generations with 95 % self-pollination and 5 % cross-pollination each generation (accumulated frequency is 5, 4 or more, 3 or more, etc.)

Number of generations of self-pollination	Pathways (32 in total)	Frequency of inbred lines	Total frequency	Accumulated frequency
Five	$(0.95)^5$	0.7738	0.7738	$0.774 = (0.95)^5$
Four	$(0.05)(0.95)^4$	0.0407	0.0407	$0.815 = (0.95)^4$
Three	$(0.95)(0.05)(0.95)^3$ $(0.05)(0.05)(0.95)^3$	0.0407 0.0021	0.0428	$0.857 = (0.95)^3$
Two	$(0.95)(0.95)(0.05)(0.95)^2$ $(0.95)(0.05)(0.05)(0.95)^2$ $(0.05)(0.95)(0.05)(0.95)^2$ $(0.05)(0.05)(0.05)(0.95)^2$	0.0407 0.0021 0.0021 0.0001	0.0451	$0.903 = (0.95)^2$
One	$(0.95)(0.95)(0.95)(0.05)(0.95)$ $(0.95)(0.95)(0.05)(0.05)(0.95)$ $(0.95)(0.05)(0.95)(0.05)(0.95)$ $(0.05)(0.95)(0.95)(0.05)(0.95)$ $(0.95)(0.05)(0.05)(0.05)(0.95)$ $(0.05)(0.95)(0.05)(0.05)(0.95)$ $(0.05)(0.05)(0.95)(0.05)(0.95)$ $(0.05)(0.05)(0.05)(0.05)(0.95)$	0.0407 0.0021 0.0021 0.0021 0.0001 0.0001 0.0001 0.0000	0.0475	$0.950 = (0.95)^1$
Zero	(16 pathways) * * * * (0.05)	0.0500	0.0500	1.000

with 5 % self pollination, this will still be largely true as the probability of any highly inbred lines will be very low; less than one in a million will be inbred for five or more generations.

### ***Sexual Reproduction with a Mixture of Self- and Cross-Pollination***

Although most crops reproduce predominantly either by self-pollination (inbreeders) or by cross-pollination (outbreeders), there are exceptions. In *Brassica napus* and *Vicia faba* (Faba or field bean) there is normally no effective mechanism to prevent self-pollination but insect pollination does result in variable amounts of outcrossing. Becker et al. (1992) reported 12–47 % outcrossing in the Swedish spring rapeseed cv ‘Topas’ over five locations in Sweden, Denmark and Germany, and on average from 11 to 39 % from top to bottom of the plant. In swedes, Gowers (1981) found 17–19 % outcrossing in three cultivars, but 46 % in a fourth cultivar as a result of the presence of some plants with high levels of self-incompatibility. Hence outcrossing rates from 20 to 40 % are often quoted for *B. napus* (Becker et al. 1992). In *Vicia faba* the degree of cross-fertilization is about 50 % (Link and

Ghaouti 2012), but highly variable, depending on the supply of pollinators, the environment, the genotype and the inbreeding status of the plant. For example, Link et al. (2005) reported values of 38–73 % for genotypic variation, with an average of 54 %, and values of 43–74 % for environmental variation, also with an average of 54 %. With a mixture of self- and cross-pollination, the composition of a landrace will be a complex intermediate between the extremes of either self- or cross-pollination. An example of the complexity can be seen in the study of Barnaud et al. (2008) on five landraces of sorghum (*Sorghum bicolor*) grown in one field by Duupa farmers in Cameroon. Although sorghum is a wind pollinated hermaphrodite that largely self-pollinates, they found that outcrossing rates varied greatly among landraces from 5 to 40 %, and that the outcrossing rate for all progenies examined was 18 % on average, but ranged from 0 to 73 % among progenies. They discuss this variation in terms of plant and inflorescence morphology and floral traits. Thus small plants, loose panicles and open flowers favour outcrossing whereas closed flowers (cleistogamy) prevent cross-pollination. Furthermore, the results showed that individual maternal plants were usually pollinated by eight to ten pollen donors, except for one landrace with a maximum of three. The outcrossing rates were multilocus, molecular marker estimates and the paper explains their superiority over single locus ones where mating between relatives (biparental inbreeding) results in some outcrossing events being confounded with selfing events. More information can be found in a paper by Ritland (2002).

### ***Sexual Reproduction Followed by Vegetative Propagation***

**Cassava:** Cassava (*Manihot esculenta*) provides us with an example where the functioning of mixed sexual/clonal systems has been studied in detail in terms of the ecology and genetics of the process. The following brief account is taken from the review by McKey et al. (2010). Cassava's starch-rich roots provide the staple food for more than 600 million people throughout the tropics. Most cultivars of cassava have retained sexual fertility, with farmers regularly incorporating 'volunteer' plants from seeds into the stock of clonal landraces which they propagate by stem cuttings. The process has been studied in the fields of traditional Amerindian farmers in Amazonia. Sexual reproduction begins when stingless bees and other insects pollinate the plant's flowers. Some self-pollination occurs as well as cross-pollination. When the fruit matures, it dries and dehisces explosively, scattering seeds on the ground up to several metres from the mother plant. Ants transport and bury the seeds in their nest or in refuse heaps nearby, thus establishing a soil seed bank in which seeds can remain dormant for up to dozens of years. They germinate when the vegetation cover is removed by a disturbance, such as field clearing and burning, which heats the soil. This happened to the wild ancestors of cassava when fire occurred in forest-savannah habitats. Seedlings thus emerge when a farmer

opens a new field by clearing and burning an old fallow and plants stem cuttings. Young plants in the field are therefore a mixture of planted clones and recombinant genotypes from sexual reproduction. Amerindian farmers can distinguish plants derived from seed from those derived from stem cuttings. Hence they can spare volunteer plants when weeding and examine those that survive to harvest. They can then incorporate some of these into their stock of clonal propagules, each usually being assigned by the farmer to the landrace it most resembles. Each landrace is thus a diverse assemblage of multiple clones sharing phenotypic characteristics. Both natural and artificial selection favour outcrossed, highly heterozygous volunteer plants, which are larger than inbred ones. Thus incorporation of the relatively few desirable new variants, and the elimination of the many undesirable ones, involves little work for the farmer who benefits from the advantages of both sexual and clonal reproduction while minimizing their disadvantages. The farms therefore combine two functions, production today and the generation of new genotypes that will ensure continued adaptation and production tomorrow. In modern farming, these activities are now commonly performed by separate populations in farmers' fields and in breeding stations, respectively. Another example of the role of sexual reproduction in the diversification of cassava is provided by the current variation found between and within landraces in the South Pacific archipelago of Vanuatu. The crop was introduced to the archipelago in the 1850s and subsequently grown by farmers in many different villages (Sardos et al. 2008).

*Potato:* The potato (*Solanum tuberosum*) provides us with another example of vegetatively propagated landraces, which nevertheless arose through sexual reproduction. Potatoes, like their ancestral wild species, reproduce by sexual means as well as by setting tubers. This sexual reproduction creates an abundance of diversity by recombining the variants of genes that arose by mutation. The genetically unique seedlings that grow from true seeds produce tubers that can be replanted as seed tubers. No doubt domestication in the Andes occurred by human selection of tubers from naturally occurring variation, and the same was true for the subsequent farmer selection and propagation of landraces of potato. Thus landrace populations are mixtures of vegetatively propagated clones, but in considering their likely genetic structure we are faced with two complications. Firstly, naturally occurring autotetraploid types of potato came to be selected in preference to their diploid ancestors, presumably because Andean farmers found them superior for yield and other traits. Today, for example, 2644 out of the 3527 landraces of potato maintained in the world collection at the International Potato Center (CIP) in Lima, Peru belong to tetraploid Group Andigena. Secondly, the gametophytic self-incompatibility of their diploid progenitors does not operate in these autotetraploids, and hence self-pollination can occur despite natural pollination by insects capable of buzz pollination. Brown (1993) estimated 40 % (range 21–74 %) natural cross-pollination in Group Andigena in the Andes whereas Glendinning (1976) found 20 % (range 14–30 %) in an artificially constructed Andigena population. Therefore in considering

the genetic structure of landraces of potato we will need to explore mixed selfing and random mating under tetrasomic inheritance.

*Yams:* My last example is from yams, important staple tuber crops in tropical and subtropical regions, in which outcrossing is encouraged by separate male and female plants. Details can be found in the review by Arnau et al. (2010), along with an account of modern breeding which started in the 1960s. The four main cultivated yams were independently domesticated on three continents some 7000 years ago: *Dioscorea rotundata* and *D. cayenensis* in West Africa, *D. alata* in Southeast Asia and the South Pacific and *D. trifida* in South America. Today, however, the major producers are in West Africa: Nigeria, Ivory Coast, Ghana and Benin. Yam fields in traditional agroecosystems are planted with tuber fragments from the previous harvest. In Benin, *D. rotundata* grows in sympatry with two wild relatives: *D. abyssinica* in northern Benin and *D. praehensilis* in the south. A practice referred to as ennoblement still occurs among farmers in Benin. They collect edible wild forms and plant them in their fields where they subject them to intense vegetative multiplication and selection over several years, leading to morphological and biochemical changes in the plants, mainly at the tuber level. The tubers are then multiplied and cultivated if farmers are satisfied with their morphology. Scarcelli et al. (2006) have provided genetic evidence that yam farmers not only select wild yams, but also spontaneous interspecific hybrids of the wild species with cultivated yams. Furthermore, they found a clear genetic signature of wild and hybrid yam in some cultivated plants. Ennoblement therefore contributes to the enhancement of the genetic diversity of cultivated landraces of yams.

### ***Concluding Remarks***

This brief survey of the composition of landraces can be summarized as follows. Depending on their mode of reproduction (sexual or asexual) and mating system (extent of self- versus cross-pollination), landraces consist of mixtures of inbred lines, or hybrids, or inbred lines and hybrids, or vegetatively propagated clones. Within these mixtures, some components will be superior to others for particular traits and a very few components will be superior for many traits. In the past both natural and human selection took place for superior components at the expense of inferior ones. The result was a great diversity of heterogeneous landraces adapted to their growing conditions and to human needs. Genetic knowledge allowed further progress to be made in the twentieth century in more efficient and sophisticated ways, and in the next chapter we shall consider population improvement leading to open-pollinated cultivars. However, plant breeding in the twentieth century was dominated by another trend, the search for the best inbred line, hybrid or clone to be

grown in monoculture over as wide an area as possible and for as many uses as possible. The result was very high yielding cultivars, but ones which were vulnerable to attack from new pests and diseases and to changes in the climate. This raises the issue of whether there really is a single best genotype in a heterogeneous population, and if not, how to identify the exceptional mixtures that out-perform their best components (Dawson and Goldringer 2012). Kiær et al. (2012) compared the yields of each of six spring barley variety mixtures with that of their three component cultivars in 17 environments, the mixtures being composed from a set of 14 cultivars. One mixture had a higher grain yield than any of its components in 8 out of 17 trials. Its average yield was  $5.37 \text{ t ha}^{-1}$  compared with 5.16 for the average of its three components and 5.21 for the highest yielding cultivar; in other words, a small 3 % advantage over the best cultivar. This demonstrates that cultivar mixtures can be found that yield as well as the best component alone, so heterogeneity does not necessarily mean a loss of yield.

In the coming chapters we will consider the breeding of inbred line, hybrid and clonal cultivars, but also the question of whether or not uniformity is always desirable. For the rest of this chapter, we will examine in a more theoretical fashion the genetic structure of populations under different mating systems.

## Mixed Selfing and Random Mating: Single Locus

Let us consider a large population of plants reproducing in non-overlapping generations in such a way that there is a constant probability  $s$  that any plant will be self-fertilized and a probability  $1 - s$  that it will cross with some plant chosen at random from the population. We will suppose that all crosses are equally fertile and all genotypes equally viable. Furthermore we will suppose that  $A$  and  $a$  are the alleles present at some locus, and their frequencies in the population are  $p$  and  $q$ . The following results are demonstrated in Box 9.1.

There are no changes in allele frequencies from generation to generation. In contrast, the frequencies of genotypes  $AA$ ,  $Aa$  and  $aa$  do change from generation to generation and asymptotically approach the following equilibrium frequencies where  $f = s/(2 - s)$  is a measure of the extent to which the population is inbred:

	$AA$	$Aa$	$aa$
From selfing	$s[p^2 + pq/(2 - s)]$	$2pq(1 - s)/(2 - s)$	$s[q^2 + pq/(2 - s)]$
From random mating	$(1 - s)p^2$	$2pq(1 - s)$	$(1 - s)q^2$
Total	$p^2 + pqs/(2 - s)$ $= p^2 + fpq$	$4pq(1 - s)/(2 - s)$ $= 2pq(1 - f)$	$q^2 + spq/(2 - s)$ $= q^2 + fpq$

**Box 9.1: Mixed Selfing and Random Mating: Single Locus**

Suppose  $A$  and  $a$  are the alleles present at some locus and their frequencies in the population are  $p$  and  $q$ . If  $P_n(AA)$  denotes the frequency (or proportion) of occurrence of the genotype  $AA$  in the  $n$ th generation, and likewise for  $Aa$  and  $aa$ , and  $s$  is the probability of self-pollination, we have:

Change in genotype frequencies			
<i>nth</i> generation	$AA$	$Aa$	$aa$
Frequency	$P_n(AA)$	$P_n(Aa)$	$P_n(aa)$
<i>(n + 1)</i> th generation			
From selfing	$s[P_n(AA) + \frac{1}{4}P_n(Aa)]$	$s[\frac{1}{2}P_n(Aa)]$	$s[\frac{1}{4}P_n(Aa) + P_n(aa)]$
From random mating	$(1-s)p^2$	$2(1-s)pq$	$(1-s)q^2$

Hence:

$$P_{n+1}(AA) = s \left[ P_n(AA) + \frac{1}{4}P_n(Aa) \right] + (1-s)p^2 \text{ and}$$

$$P_{n+1}(Aa) = \frac{1}{2}sP_n(Aa) + 2(1-s)pq \text{ and}$$

$$P_{n+1}(aa) = s \left[ P_n(aa) + \frac{1}{4}P_n(Aa) \right] + (1-s)q^2$$

No change in allele frequency

The frequency of allele  $A$  in generation  $n$  is:

$$p = P_n(AA) + \frac{1}{2}P_n(Aa)$$

The frequency of allele  $A$  in generation  $(n + 1)$  is:

$$\begin{aligned} P_{n+1}(AA) + \frac{1}{2}P_{n+1}(Aa) &= s \left[ P_n(AA) + \frac{1}{2}P_n(Aa) \right] + (1-s)p^2 \\ &\quad + (1-s)p(1-p) \\ &= sp + (1-s)p = p \end{aligned}$$

Hence there is no change in allele frequency from generation to generation.

Equilibrium genotype frequencies

Let  $K_n = P_n(Aa) - 4pq(1-s)/(2-s)$

(continued)

**Box 9.1** (continued)

Then

$$\begin{aligned}
 K_1 &= P_1(Aa) - 4pq(1-s)/(2-s) = \frac{1}{2}sP_0(Aa) + 2(1-s)pq \\
 &\quad - 4pq(1-s)/(2-s) \\
 &= \frac{1}{2}sP_0(Aa) + 2pq(1-s)[1 - 2/(2-s)] = \frac{1}{2}sP_0(Aa) \\
 &\quad - 2pq(1-s)s/(2-s) \\
 &= \frac{1}{2}s[P_0(Aa) - 4pq(1-s)/(2-s)] = \frac{1}{2}sK_0
 \end{aligned}$$

Hence  $K_n = (\frac{1}{2}s)^n K_0$

As  $n$  gets larger and larger, because  $s \leq 1$ ,  $\left(\frac{1}{2}s\right)^n$  gets smaller and smaller, so that in the limit of an infinite number ( $\infty$ ) of generations,  $K$  becomes zero, and we have the equilibrium frequency:

$$P_\infty(Aa) = 4pq(1-s)/(2-s) = 2pq - 2fpq,$$

where  $f = s/(2-s)$  is a measure of extent to which population is inbred. Likewise:

If  $K_n = P_n(AA) - (p^2 + pqs/(2-s))$ , then  $K_n = \left(\frac{1}{2}s\right)^n K_0$ , and

$$P_\infty(AA) = p^2 + pqs/(2-s) = p^2 + fpq$$

If  $K_n = P_n(aa) - (q^2 + pqs/(2-s))$ , then  $K_n = \left(\frac{1}{2}s\right)^n K_0$ , and

$$P_\infty(aa) = q^2 + pqs/(2-s) = q^2 + fpq$$

### Rate of approach to equilibrium

$$K_n = \left(\frac{1}{2}s\right)^n K_0$$

Hence  $\ln K_n = n \ln \left(\frac{1}{2}s\right) + \ln K_0$

And the graph of  $\ln K_n$  against  $n$  is a straight line of slope  $\ln \left(\frac{1}{2}s\right)$  which is negative as  $s \leq 1$ .

**Table 9.2** Genotype frequencies at a single locus under mixed selfing and random mating where  $s$  is the probability of self-pollination

$s$	$AA$	$Aa$	$aa$
0	$p^2$	$2pq$	$q^2$
0.05	$p^2 + 0.03pq$	$2pq(0.97)$	$q^2 + 0.03pq$
0.2	$p^2 + 0.11pq$	$2pq(0.89)$	$q^2 + 0.11pq$
0.3	$p^2 + 0.18pq$	$2pq(0.82)$	$q^2 + 0.18pq$
0.7	$p^2 + 0.54pq$	$2pq(0.46)$	$q^2 + 0.54pq$
0.8	$p^2 + 0.67pq$	$2pq(0.33)$	$q^2 + 0.67pq$
0.95	$p^2 + 0.90pq$	$2pq(0.10)$	$q^2 + 0.90pq$
1	$p$	0	$q$

Some numerical results are given in Table 9.2. The frequency of heterozygotes ( $2pq$ ) is at a maximum when  $p = q = \frac{1}{2}$ , as can be demonstrated by considering the derivative of  $2pq = 2q(1 - q)$  with respect to  $q$  (it is  $2(1 - 2q)$ , which is a maximum when  $q = \frac{1}{2}$ ). When  $s = 0$  (no selfing), we have an outbreeding population in Hardy-Weinberg equilibrium. In 1908, Godfrey Harold Hardy and Wilhelm Weinberg independently realized that the consequence of Mendelian inheritance in a large random-mating population with no selection, mutation or migration was that there were no changes in the gene frequencies and the genotype frequencies from generation to generation. In other words, the mechanism of inheritance maintains variation. Furthermore, Hardy and Weinberg derived the simple relationship between gene and genotype frequencies which became known as the Hardy-Weinberg law. One generation of random mating establishes the equilibrium. Strictly speaking, the heredity mechanism maintains allele frequencies and the mating system determines how gametes are combined into zygotes and thus zygote frequencies. Once some self-pollination takes place, the new equilibrium is only approached asymptotically, but in practice, a population will not be far from equilibrium after three to five generations. When  $s = 1$  (complete selfing), the population is a mixture of homozygotes in the ratio of  $p$  AA to  $q$  aa.

In order to get a better feel for the genetic structure of populations we need to consider two loci. But first, there is an important point to make about the effect of past selection, because landraces will have been subjected to both farmer selection and natural selection. Such selection not only alters allele frequencies but also generates disequilibrium, as can be seen in the following simple example.

## Mixed Selfing and Random Mating: Selection

Let's consider a population in which the frequency of allele  $A$  equals the frequency of  $a$  (both  $\frac{1}{2}$ ) and the proportion of selfing  $s$  is a half. Furthermore, let us consider both a simple additive model in which the phenotypes of AA, Aa

and  $aa$  are 1, 0 and  $-1$ , and one with complete dominance in which the phenotypes of  $AA$ ,  $Aa$  and  $aa$  are 1, 1 and  $-1$ . The equilibrium structure and mean of the population is as follows:

	$AA$	$Aa$	$aa$
Frequency	1/3	1/3	1/3

Population mean = 0 and  $1/3$  for additive and complete dominance model.

If we create a new population by rejecting all of the individuals with a phenotypic score of  $-1$ , that is  $AA$  and  $Aa$  genotypes are selected and  $aa$  genotypes are discarded, and then allow mixed selfing and random mating to occur, the frequency of  $A$  is now  $3/4$ , the frequency of  $a$  is  $1/4$ , and the structure and mean of the population are as follows:

	$AA$	$Aa$	$aa$
Frequency	19/32	10/32	3/32

Population mean =  $1/2$  and  $13/16$  for additive and dominance model.

Hence we have increased the frequency of allele  $A$  and the population mean, but this is not the equilibrium structure of the population which will be achieved by further generations of mixed selfing and random mating, but without selection:

	$AA$	$Aa$	$aa$
Frequency	20/32	8/32	4/32

Population mean =  $1/2$  and  $12/16$  for additive and dominance model.

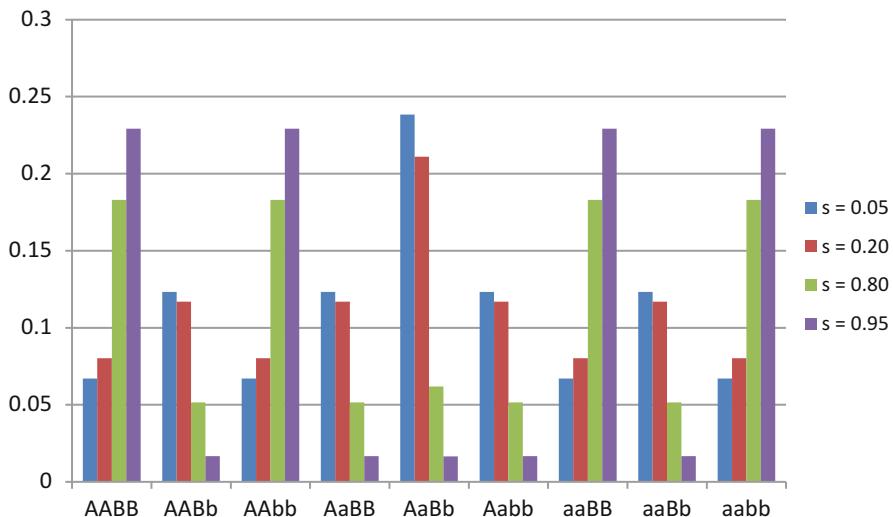
Hence in the presence of dominance, the population mean can fall as the new landrace is multiplied without selection. This is a simple example of selection with partial selfing which was considered by Wright and Cockerham (1985, 1986) for a model with arbitrary additive and dominance effects. They demonstrated that the response measured in the first generation offspring after selection (immediate gain) can differ from that retained when the population has regained equilibrium (permanent effect), especially with the large inbreeding depression associated with increasing dominance levels and numbers of loci.

## Mixed Selfing and Random Mating: Two Loci

It is tempting but wrong to think that the equilibrium for two loci is simply the single locus ones multiplied together. The correct derivation can be found in Bennett and Binet (1956). For two unlinked loci, however, there is a simple correction that can be added or subtracted, as shown in Table 9.3. For given gene frequencies, this correction has its greatest value of  $0.1506pquv$  when  $s = 0.6946$  ( $p$ ,  $q$ ,  $u$  and  $v$  are the frequencies of alleles  $A$ ,  $a$ ,  $B$  and  $b$ ). When  $s = 0$  or  $s = 1$ , there is no correction.

**Table 9.3** Correction for equilibrium frequencies of two unlinked loci under mixed selfing and random mating: single locus frequencies are multiplied together and then corrected, where  $x = 4s(1-s)pquv/(4-s)(2-s)^2$ ,  $s$  is the proportion of selfing and  $p, q, u$  and  $v$  are the frequencies of alleles  $A, a, B$  and  $b$

Genotype	Frequency	Single loci	Correction
$AABB$	$P(AABB)$	$P(AA)P(BB)$	$x$
$AABb$	$P(AABb)$	$P(AA)P(Bb)$	$-2x$
$AAAb$	$P(AAbb)$	$P(AA)P(bb)$	$x$
$AaBB$	$P(AaBB)$	$P(Aa)P(BB)$	$-2x$
$AaBb$	$P(AaBb)$	$P(Aa)P(Bb)$	$4x$
$Aabb$	$P(Aabb)$	$P(Aa)P(bb)$	$-2x$
$aaBB$	$P(aaBB)$	$P(aa)P(BB)$	$x$
$aaBb$	$P(aaBb)$	$P(aa)P(Bb)$	$-2x$
$aabb$	$P(aabb)$	$P(aa)P(bb)$	$x$



**Fig. 9.2** Genotype frequencies for the situation when  $p = q = u = v = \frac{1}{2}$  (equal allele frequencies at both loci) and  $s = 0.05, 0.20, 0.80$  and  $0.95$  (proportion of selfing)

Figure 9.2 shows the difference in genetic makeup of predominantly crossbreeding ( $s = 0.05$  and  $0.20$ ) and inbreeding ( $s = 0.80$  and  $0.95$ ) populations. The latter is dominated by homozygosity at both loci with a low level of heterozygosity at one or both loci. If there was no crossbreeding, then there would be no heterozygosity.

Gale (1980) in his book *Population Genetics* explains how one can estimate the proportion of lines homozygous at all loci, without knowing the number of loci. The information required is the haploid chromosome number ( $v$ ) of the plant species and its total map length ( $L$ ) defined as half the number of chiasmata per nucleus, and measured in Morgans, together with an understanding of meiosis. It is assumed that

at the beginning every locus is heterozygous. Then after  $n$  generations, the proportion of lines homozygous at all loci is  $e^{-m}$ , where  $m$  is the number of heterozygous tracts (stretches of chromosome that are not homozygous) averaged over lines.

$$m = L \left( \frac{1}{2} \right)^n [2n + v/L]$$

The conclusion from Gale's example of *Papaver dubium* is that it takes about 16 generations for homozygosity at all loci to be a near certainty, whereas after 8 generations only 0.39 % of lines are still heterozygous at any given locus, and the total length of heterozygous tracts is quite small in virtually all lines. In a practical plant breeding programme, as we shall see later, complete homozygosity can be quickly achieved through the production of doubled haploids. But the reason for doing so is to produce lines for assessment that are true breeding for important traits as quickly as possible, rather than the achievement of complete homozygosity.

Bennett and Binet (1956) also showed that when two loci are linked on the same chromosome, the equilibrium frequencies of genotypes under mixed selfing and random mating depend on the recombination frequency  $r$  between loci, as well as on  $s$ . For example:

$$P_\infty(AB/ab) = P_\infty(Ab/aB) = 4(1-s)pquv/[2 - s + 2sr(1-r)]$$

We now have to distinguish  $A$  and  $B$  in coupling ( $AB$  and  $ab$  linkage) from  $A$  and  $B$  in repulsion ( $Ab$  and  $aB$  linkage) rather than simply write  $AaBb$  for unlinked loci.

When  $r = 0.5$ :

$$P_\infty(AB/ab) = P_\infty(Ab/aB) = 4(1-s)pquv/[2 - 0.50s]$$

When  $r = 0.1$ :

$$P_\infty(AB/ab) = P_\infty(Ab/aB) = 4(1-s)pquv/[2 - 0.82s]$$

In other words, the frequency of the double heterozygotes at equilibrium increases with tighter linkage.

## Random Mating: Two Loci

It is worth considering the special case of no selfing. When  $s = 0$ , from the previous section we have at equilibrium:

$$P_\infty(AB/ab) = P_\infty(Ab/aB) = 2pquv$$

At equilibrium the frequencies of gametes can be determined by multiplying together the frequencies of the alleles in the gametes:

$$P(AB) = pu, P(Ab) = pv, P(aB) = qu \text{ and } P(ab) = qv,$$

where  $P(AB)$ ,  $P(Ab)$ ,  $P(aB)$  and  $P(ab)$  are the frequencies of the gametes  $AB$ ,  $Ab$ ,  $aB$  and  $ab$ .

Hence  $P(AB)P(ab) = P(Ab)P(aB) = pquv$ .

Then from the random union of gametes we have  $P(AB|ab) = 2P(AB)P(ab) = 2pquv$ .

This equilibrium is called linkage equilibrium, and is achieved asymptotically over generations ( $n$ ), but is approached after five generations of random mating when  $r = 0.5$ . However, very tight linkage can maintain disequilibrium over many generations, and this is the basis of association mapping which we encountered in Chap. 6. The interested reader will find the derivation of the following results in any standard textbook of Population Genetics (for example, *Genetics of Populations, Fourth Edition* by Hedrick (2011)).

The amount of disequilibrium  $D$  is measured as follows:

$P(AB) = pu + D$ ,  $P(Ab) = pv - D$ ,  $P(aB) = qu - D$  and  $P(ab) = qv + D$ , where the frequencies of alleles  $A$ ,  $a$ ,  $B$  and  $b$  are  $p$ ,  $q$ ,  $u$  and  $v$ .

It can be shown that  $D = P(AB)P(ab) - P(Ab)P(aB)$ , so that the range of values of  $D$  is 0.25 (when  $P(AB) = P(ab) = \frac{1}{2}$ ,  $P(Ab) = P(aB) = 0$ ) to -0.25 (when  $P(AB) = P(ab) = 0$ ,  $P(Ab) = P(aB) = \frac{1}{2}$ ).

By considering the change in gamete frequencies from one generation to the next for two linked loci each with two alleles, it can also be shown that  $D_1 = (1 - r)D_0$  and hence

$$D_n = (1 - r)^n D_0.$$

And the number of generations ( $n$ ) for  $D_0$  to decay to  $D_n$  is

$$n = \ln(D_n/D_0)/\ln(1 - r).$$

Although  $D$  has a number of good algebraic properties as a measure of linkage disequilibrium (LD), it varies in size as a function of the frequencies of its constituent alleles. Hence other measures of LD are sometimes used:

$$\begin{aligned} D' &= D/D_{\max} (\text{range } -1.0 \text{ to } 1.0, \text{ regardless of allele frequency}) \\ r^2 &= D^2/(pquv) (\text{range } 0 \text{ to } 1.0, \text{ but depends on allele frequencies}) \end{aligned}$$

If  $r' = r/r_{\max}$ , then  $r' = D'$  (see Hedrick 2011).

Selection can generate disequilibrium, as we shall see in the next chapter. Hence the genetic structure of a landrace that has undergone recent selection is expected to change on multiplication without selection. If non-allelic interactions (epistasis) are present, the population mean can fall during this multiplication.

## Linkage Disequilibrium Under Mixed Selfing and Random Mating

Weir and Cockerham (1973) derived the asymptotic decay rate of linkage disequilibrium in terms of the proportion of selfing ( $s$ ) and recombination ( $r$ ) as:

$$D_{n+1} = \frac{1}{2} \left\{ (1 + \lambda + s)/2 + [((1 + \lambda + s)/2)^2 - 2s\lambda]^{\frac{1}{2}} \right\} D_n,$$

where  $\lambda = 1 - 2r$ , so that the proportion of selfing and recombination are on the same scale.

When  $s = 0$  (random mating):

$D_{n+1} = \frac{1}{2} \left\{ (1 + \lambda)/2 + (1 + \lambda)/2 \right\} D_n = \frac{1}{2}(1 + \lambda)D_n = (1 - r)D_n$ , as given in the previous section.

If the two genes are unlinked ( $r = 0.5$  and  $\lambda = 0$ ),  $D_{n+1} = \frac{1}{2}(1 + s)D_n$ . In other words,  $s$  has the same effect as  $\lambda$ . However, if both partial selfing and linkage are present, they can have a marked effect on the rate of decay, bringing it nearly to a standstill. When  $s = 0.99$  and  $r = 0.01$ , it takes 3570 generations for the equilibrium to be halved (Hedrick 2011). In Chap. 13 on inbred line cultivars we will come across the situation of making crosses to generate genetic variation and then using 100 % self-pollination to return to homozygosity. Under these circumstances, despite some recombination, the rapid approach to homozygosity under recurrent selfing prevents linkage equilibrium ever being attained among linked genes.

## Mixed Selfing and Random Mating: Tetrasomic Inheritance

As explained earlier, in considering the genetic structure of landraces of potato we need to explore mixed selfing and random mating under tetrasomic inheritance. The equilibrium for two alleles,  $A$  and  $a$ , in a tetraploid population which is partly self-pollinated was derived by Haldane (1930). The algebra is quite straightforward but very tedious and hence will not be repeated here. One starts with the five genotypes and their frequencies at equilibrium and then derives the five genotype frequencies in the next generation, based on a constant probability  $s$  that any plant will be self-fertilised and a probability  $1 - s$  that it will cross with some plant chosen at random from the population. As there is no change in frequency from one generation to the next at equilibrium, the relevant equations can be solved to express the genotype frequencies in terms of  $s$  and the ratio of the frequency of  $A$  ( $p$ ) to  $a$  ( $q$ ). The results for  $s = 0.2, 0.4, 0.6$  and  $0.8$  are given in Table 9.4 for  $p = q = \frac{1}{2}$ . It can be seen that

with 40 % ( $s=0.6$ ) cross-pollination the  $AAaa$  genotype is most frequent whereas with 20 % ( $s=0.8$ ) cross-pollination  $AAAA$  and  $aaaa$  are the most frequent genotypes, followed by  $AAaa$ .

**Table 9.4** Genotype frequencies at equilibrium for a single locus with alleles  $A$  and  $a$  at frequencies  $p=q=\frac{1}{2}$  under mixed selfing and random mating with tetrasomic inheritance, ignoring the phenomenon of double reduction, where  $s$  is the proportion of selfing

Genotype	$s = 0$	$s = 0.2$	$s = 0.4$	$s = 0.6$	$s = 0.8$	$s = 1$
$AAAA$	$p^4 = 0.0625$	0.0874	0.1239	0.1814	0.28245	$p = 0.5$
$AAAa$	$4p^3q = 0.25$	0.23865	0.21865	0.1836	0.1202	0
$AAaa$	$6p^2q^2 = 0.375$	0.3479	0.3149	0.2700	0.1947	0
$Aaaa$	$4pq^3 = 0.25$	0.23865	0.21865	0.1836	0.1202	0
$aaaa$	$q^4 = 0.0625$	0.0874	0.1239	0.1814	0.28245	$q = 0.5$

## Concluding Remarks

The theoretical results have confirmed that predominantly self-pollinating populations will be mixtures of inbred lines with a high degree of homozygosity, whereas predominantly cross-pollinating ones will be mixtures of hybrids with a high level of heterozygosity. They have also revealed the complexity that can be expected with a mixture of self- and cross-pollination in both diploid and autotetraploid species. Furthermore, they have demonstrated that the genetic structure of a population is expected to change over generations until a dynamic equilibrium is reached, with high levels of self-pollination and tight linkage between loci causing a slow approach to equilibrium. However, in landraces both human and natural selection will also be operating over generations and there may be some human exchange of seed from neighbouring landraces, as well as some pollen flow in outcrossing species. Hence landraces should be viewed as evolving entities, in contrast to modern cultivars which are expected to be maintained true to type (the definitive stock), and where seed production systems to achieve this are a very important final stage in breeding.

# **Chapter 10**

## **Open-Pollinated and Synthetic Cultivars from Population Improvement**

### **Introduction**

All plant breeding comprises recurrent cycles of crossing followed by selection, as we shall see in this and subsequent chapters. Here we are going to explore how the application of Mendelian genetics in the twentieth century allowed various methods of population improvement to be developed for the production of new open-pollinated and synthetic cultivars. We will consider the factors that need to be taken into account when designing such a population improvement programme.

### **Landraces to Open-Pollinated Cultivars**

The nineteenth century saw the start of the transition from landraces to modern cultivars. Maize in Central and North America provides us with as good an example as any crop of the changes from domestication, through landraces to open-pollinated cultivars. Summaries can be found in chapter 40 (by Goodman) of *Evolution of Crop Plants, Second Edition*, edited by Smartt and Simmonds (1995), chapter 9 of *Principles of Plant Breeding, Second Edition* by Allard (1999) and in the papers by Smith (1989) and Troyer (2006). Some of the details are still the subject of research and debate, particularly the origin of the ‘northern flints’. Briefly some of the key points are as follows.

### ***Maize***

Maize was domesticated in the mid- to lowland regions of southwest Mexico around 9000 years ago (van Heerwaarden et al. 2011) and subsequently introduced into North America in areas east of the Mississippi River some 2000 years ago.

There the agriculture of the eastern woodlands was already well developed, being based on several endemic starchy and oily small-seeded annuals as well as nuts from several perennial trees. These had been domesticated between 4000 and 3000 years ago. Then from 1200 to 900 years ago, maize production rapidly expanded and largely replaced the old endemic cropping system in the eastern woodlands, from North Florida to the northern limits of maize growing. By then, two types of maize had evolved and were grown, one adapted to the warm southeast and the other to the cooler mean temperatures and longer days, but shorter growing seasons of Canada and central North America. Northern flint maize rapidly gained ascendancy in the northern and central regions of North America. By the time Columbus discovered the New World in 1492, over 200 distinctive varieties of flint maize had been developed by North American Indian farmers. About this time higher-yielding white dent maize started to be grown in northeastern Mexico and soon after in coastal southeastern North America, but stayed separate from the flint type. Flint maize has a mostly soft, opaque, floury endosperm in contrast to dent maize which has dented kernels as a result of a central core of soft starch shrinking more in drying than the surrounding hard endosperm. The emergence, recognition and deliberate maintenance by farmers of distinctive varieties of maize marked the transition from landraces to open-pollinated cultivars. The open-pollinated, common, yellow dent cultivars of the United States Corn Belt were the creation of American farmers and plant breeders during the nineteenth century.

The period from 1800 onwards saw a rapid westward and northward migration of maize in the USA, from Tennessee to Illinois, Iowa and Minnesota, and the selection of varieties adapted to these different growing conditions. More than 500 open-pollinated cultivars of maize are thought to have existed in the United States and Canada by 1840, and more than 1000 by 1920. One very popular open-pollinated maize cultivar over much of the USA was Reid Yellow Dent, but serendipity played a part in its origin. When Robert Reid moved from Ohio to Illinois in 1846, he late-planted a dent maize (a Virginian gourdseed dent known as Gordon Hopkins) and hence harvested a crop with many immature kernels, seed of which produced a poor stand the following year. He replanted the gaps with an early maturing, native Indian yellow flint maize (Little Yellow). The two types cross pollinated; and over 40 subsequent years and generations Robert Reid, and then his son James, practised selection for what they considered to be desirable ears and kernels from plants with high yield and appropriate maturity. James Reid also started a seed business to supply farmers with seed of Reid Yellow Dent. As a result many local strains were developed which differed in their yielding power, although they could be visually recognized as Reid's Yellow Dent. Two other cultivars that later became important in hybrid development (Chap. 12) were 'Lancaster Sure Crop' from Pennsylvania and 'Krug' from Illinois. Many nineteenth century maize farmers did make deliberate crosses between cultivars, and as mentioned earlier, it was William Beal (1833–1910) of Michigan Agricultural College who first published results on crossing maize cultivars in 1876.

Nevertheless, as we saw in Chap. 2, maize yields in the USA were practically static from 1865 to 1927. Had population improvement methods been available and applied to maize in the USA at the end of the nineteenth century, yields would probably have risen. However, in Chap. 12 we shall see how open-pollinated cultivars were used as the source of inbred lines that were then crossed to produce hybrid cultivars, with immediate and dramatic increases in yield. As a consequence, there was initially less emphasis on long term population improvement.

### ***Perennial Forage Crops and Amenity Grasses***

Although much research has been done with maize, population improvement leading to an open-pollinated cultivar is a natural choice of breeding programme for any outbreeding seed propagated crop. It has proved popular in perennial forage crops and amenity grasses over the last 100 and last 50 years, respectively (Posselt 2010). In these crops diverse natural ecotypes (populations adapted to different ecological conditions), rather than landraces, were commonly used during the twentieth century to start scientific breeding programmes (Boller et al. 2010a). In perennial forage crops and amenity grasses we also see another major development, the production of synthetic cultivars from improved populations.

## **Foundation Population**

At the beginning of any breeding, objectives are determined followed by decisions on the choice of germplasm for starting the programme. Wrong objectives and wrong germplasm will result in failure many years later. The long term selection experiment in maize described at the end of this chapter started with a single open-pollinated cultivar. More usually a number of open-pollinated cultivars will have been assessed and high yielding, adapted ones with complementary traits included in the foundation population. Iowa Stiff Stalk Synthetic (BSSS) maize was developed in 1933/1934 from 16 inbred lines which had been selected by various breeders for resistance to stalk breakage from open-pollinated cultivars. It has been used extensively in selection programmes for yield improvement and resistance to maize pests (Hallauer et al. 2010).

A non maize example is the breeding programme I started in 1981 to improve fodder kale (*Brassica oleracea*), which is fed to cattle and sheep in autumn and winter in the United Kingdom and New Zealand. As clubroot (*Plasmodiophora brassicae*) is the most serious disease of kale and other Brassicas, 120 resistant plants were chosen from the 16 most resistant cultivars out of 96 assessed. Genetic variation was achieved by including cultivars of diverse geographic origin, but

restricted to the marrow-stem type for eventual ease of producing a cultivar with the necessary uniformity for National Listing (Bradshaw and Wilson 2012). In perennial forage and amenity grasses, as mentioned above, diverse natural ecotypes were commonly used during the twentieth century to start scientific breeding programmes (Boller et al. 2010a). Today diversity among and within populations can be measured using molecular markers as well as morphological characters and any known pedigree information. It is however important to check the overall performance of the new population, as ones with a large phenotypic variance often have lower means than those with a narrow base. One does not want a new population to be too inferior to the best cultivars available or it could be a number of generations before the new population becomes competitive.

Ideally sufficient generations of random mating (either natural or artificial pollination) should follow to allow all possible combinations of new traits to occur; in other words, two generations with four parents, three with eight, four with 16, etc. In a vegetable breeding programme started in 1997 at Oregon State University to develop an open-pollinated broccoli (*Brassica oleracea*) cultivar for organic production, 4 years (generations) of random mating without selection took place to maximize recombination within the population, followed by 7 years of selection with farmer participation before new cultivars were sought (Myers et al. 2012). In other situations there may be financial pressure to start selection immediately.

## Population Improvement

Although virtually any population improvement scheme is possible in maize, in other crops their reproductive biology may impose constraints on which ones are feasible in practice. Within these practical constraints, a breeder will want to choose the best scheme possible, but this may involve decisions about the balance between rapid short term gains and greater overall gains in the medium to long term. In this section we will look at the guidance available from simple theory, the genetic analysis required to make predictions, and the need to complement these with computer simulations of more realistic and hence complicated situations.

With natural cross-pollination, the seeds collected from individual plants will be half-sib families, provided self-pollination was prevented (naturally or by emasculation, for example) and pollen came from more than one other plant. Simple selection theory assumes that the pollen was a random sample from many other plants. The seed can be bulked (equal quantities from each plant) before sowing, assessing and then selecting individual (spaced) plants for further multiplication (mass selection). Alternatively, the seeds from individual plants can be kept separate and whole half-sib families sown (as spaced plants or at normal planting density) and assessed before selecting the best for further multiplication (half-sib

family selection). Ear-to-row selection in maize and maternal line selection in forage crops (alfalfa, clover and grasses) is in essence half-sib family selection, and modified ear-to-row selection is selection among and within half-sib families. Casler and Brummer (2008) have discussed the theoretical advantages of among-and-within-family (AWF) selection methods in perennial forage crops. For traits like yield in maize, random mating will have already taken place before selections are made. In forage and vegetable crops, selections are made before the plants flower and hence both male and female parents are selected. In grasses, for example, selected plants can remain in the breeding nursery and other undesirable plants destroyed before flowering, or they can be removed and planted together in isolation. Planting schemes have been devised to encourage a homogeneous, common pollen cloud for half-sib family production; for example the polycross in which plants (or clonal propagules when feasible) are surrounded by unrelated ones as far as is possible (Posselt 2010).

Full-sib families can be produced by artificial pollination of one plant with pollen from another plant, or by natural pollination when pairs of self-incompatible plants can be isolated from other sources of pollen. In grasses, for example, large transparent bags are used to enclose pairs of plants and a strong air current is drawn through the bags to ensure mutual pollination. In maize, plants can readily be self-pollinated to produce plenty of seed of  $S_1$  families for assessment. In contrast, vegetable brassicas have a strong sporophytic self-incompatibility system and self-pollination requires bud pollination; hence the quantity of seed produced can be limiting. Grasses and leguminous forages have a gametophytic incompatibility system controlled by two loci in the ryegrasses and fescues and a single locus in the leguminous forages. Seed set after enforced selfing is largely genotype dependent in grasses, and requires tripping of flowers by hand in leguminous forages (Posselt 2010).

Where self-pollination is straightforward, selfed families can be produced in addition to the half-sib or full-sib families for assessment, and selected individuals can also be selfed. The selfed seed of the selected plants or families can then be used to produce the next generation. There is a subtle difference between using the assessed family itself rather than remnant seed of the assessed family to produce the next generation. The former, but not the latter, allows one to exploit any additional superiority of the family due to chance variation in the sample of seed which is assessed. We will also see in the next chapter on vegetatively propagated crops that many of the breeding schemes are in fact simple mass selection, where the selection is multtrait and multistage over clonal generations. Mass selection in grass breeding can also involve clonal rows as grasses can be vegetatively propagated by separation of tillers.

It should be clear from the deliberations in the previous paragraphs that it is important to clearly draw out the selection scheme under consideration before working out how to predict the response to selection. This should include whether or not remnant seed is available. If so, remnant seed of all families could be sown as

spaced plants at the same time as families are assessed in plots with normal planting density. Furthermore, within family selection could be practised on the spaced plants of selected families. Formulae and their derivations can be found in books such as *Quantitative Genetics in Maize Breeding, Third edition* (Hallauer et al. 2010). In the next section just four schemes are considered: mass selection and half-sib, full-sib and S<sub>1</sub> family selection where selection takes place before flowering and residual seed of selected families is used to produce the next generation. As some forage crops are autotetraploids (alfalfa, red clover and a number of grasses, e.g. most cultivars of cocksfoot *Dactylis glomerata* spp. *glomerata* ( $2n=4x=28$ )), mass selection and half-sib family selection under tetrasomic inheritance will also be considered.

## Mass, Half-Sib Family, Full-Sib Family and S<sub>1</sub>Family Selection

### Theory Under Disomic Inheritance

The breeding cycle comprises selection before flowering of individuals or families with desirable characteristics followed by pollination and seed production from these individuals or families (using remnant seed) to create a new population. The first stage results in changes in allele frequencies which the second stage translates into changes in genotype frequencies determined by the mating system. Simple selection theory considers the extent to which phenotypic selection changes allele frequencies followed by the effect of these changes on the population mean (change from one generation to next), assuming random-mating to produce the new population, and hence Hardy-Weinberg equilibrium. Furthermore, it is assumed that there is no correlation between the genotypes assessed and their environment (i.e. assessment is done in a properly randomized trial). Under these circumstances the covariance of allele frequencies and phenotypic values equals the covariance of allele frequencies and genotypic values, so that the regression of allele frequencies ( $\beta$ ) on phenotypic values equals the covariance of allele frequencies and genotypic values divided by the phenotypic variance. These variances, covariances and regression coefficients are calculated in Box 10.1 for mass selection (MASS) and half-sib (HS) family, full-sib (FS) family and S<sub>1</sub> (one generation of selfing) family selection. The symbol  $\alpha$  [ $= a + d(q - p)$ ] is used for more than just convenience. It is the average effect of the allele substitution ‘ $a$ ’ to ‘ $A$ ’, when ‘ $a$ ’ alleles are chosen at random from a population in Hardy-Weinberg equilibrium, as explained by Falconer and Mackay (1996). It is also the average effect of allele ‘ $A$ ’ minus the average effect of allele ‘ $a$ ’. The average effect of a particular allele is the mean deviation from the population mean of individuals which received that allele from

one parent, the allele received from the other parent having come at random from the population. The concept of average effect of an allele allows the theory to be extended from two alleles to multiple alleles, but need not concern us here. The average effect of an allele ('A') substitution is also the linear regression of genotypic value on the number of ('A') alleles in the genotype. The presence of epistasis changes the expression for  $\alpha$ , as we shall see later. The equivalent expression for S<sub>1</sub> families is  $\alpha_S = a + \frac{1}{2}d(q - p)$ . The responses to selection are predicted as follows.

**Box 10.1: Derivation of Regression ( $\beta$ ) of Allele Frequency on Phenotypic Value**

Single locus with frequencies of alleles 'A' and 'a' equal to  $p$  and  $q$ , respectively ( $p + q = 1$ )

(a) Mass (MASS) selection

Genotypes of individuals	AA	Aa	aa
Genotype frequency	$p^2$	$2pq$	$q^2$
Frequency of A	1	$\frac{1}{2}$	0
Genotypic value (or $m + a$ etc)	$a$	$d$	$-a$

Mean of individuals = population mean =  $a(p - q) + 2dpq$

Variance of individuals =  $V_G = 2pqa^2 + (2pqd)^2 = V_A + V_D$ , where  $V_A$  is the additive genetic variance,  $V_D$  is the dominance variance and  $\alpha = a + d(q - p)$ .

Covariance of allele 'A' frequency and phenotypic value equals covariance of allele 'A' frequency and genotypic value =  $pq\alpha$ .

Regression ( $\beta$ ) of allele 'A' frequency on phenotypic value:

$\beta = pq\alpha/V_P$ , where  $V_P$  is the variance of phenotypic values.

$V_P = V_G + V'_E = V_A + V_D + V'_E$ , where  $V'_E$  is the environmental component of the plant to plant variation.

(b) Half-sib (HS) family selection

Genotypes of female parent	AA	Aa	aa
Genotype frequency	$p^2$	$2pq$	$q^2$
<hr/>			
Pollen genotype A frequency $p$	AA	AA Aa	Aa
Pollen genotype a frequency $q$	Aa	Aa aa	aa
<hr/>			
Frequency of A in HS family	$p + \frac{1}{2}q$	$\frac{3}{4}p + \frac{1}{4}q$	$\frac{1}{2}p$
HS family mean	$pa + qd$	$\frac{1}{2}[a(p - q) + d]$	$-qa + pd$

(continued)

**Box 10.1** (continued)

Mean of HS families = population mean =  $a(p - q) + 2dpq$

$$\text{Variance of HS families} = V_{G(HS)} = \frac{1}{2} pqa^2 = \frac{1}{4} V_A$$

Covariance of gene frequency and phenotypic value equals covariance of gene frequency and genotypic value =  $\frac{1}{4} pqa$

Regression ( $\beta$ ) of gene frequency on phenotypic value:

$$\beta = \frac{1}{4} pqa/V_{P(HS)}, \text{ where } V_{P(HS)} \text{ is the variance of phenotypic family means.}$$

$V_{P(HS)} = V_{G(HS)} + V_E/r = \frac{1}{4} V_A + V_E/r$ , where  $V_E$  is the plot to plot environmental variation in the assessment trial and  $r$  is the number of replicates of the HS families in the trial.

(c) Full-sib (FS) family selection

Genotypes of female parent	AA	AA	AA	Aa	Aa	aa
Genotypes of male parent	AA	Aa	aa	Aa	aa	aa
Genotype frequency	$p^4$	$4p^3q$	$2p^2q^2$	$4p^2q^2$	$4pq^3$	$q^4$
<hr/>						
Genotypes in FS families	AA	$\frac{1}{2}AA$	Aa	$\frac{1}{4}AA$	$\frac{1}{2}Aa$	aa
		$\frac{1}{2}Aa$		$\frac{1}{2}Aa$	$\frac{1}{2}aa$	
				$\frac{1}{4}aa$		
<hr/>						
Frequency of A in FS family	1	$\frac{3}{4}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{4}$	0
FS family mean	$a$	$\frac{1}{2}(a+d)$	$d$	$\frac{1}{2}d$	$\frac{1}{2}(-a+d)$	$-a$

Mean of FS families = population mean =  $a(p - q) + 2dpq$

$$\text{Variance of FS families} = V_{G(FS)} = pqa^2 + (pqd)^2 = \frac{1}{2} V_A + \frac{1}{4} V_D$$

Covariance of gene frequency and phenotypic value equals covariance of gene frequency and genotypic value =  $\frac{1}{2} pqa$ .

Regression ( $\beta$ ) of gene frequency on phenotypic value:

$$\beta = \frac{1}{2} pqa/V_{P(FS)}, \text{ where } V_{P(FS)} \text{ is the variance of phenotypic family means.}$$

$V_{P(FS)} = V_{G(FS)} + V_E/r = \frac{1}{2} V_A + \frac{1}{4} V_D + V_E/r$ , where  $V_E$  is the plot to plot environmental variation in the assessment trial and  $r$  is the number of replicates of the FS families in the trial.

(continued)

**Box 10.1** (continued)(d) S<sub>1</sub> family selection

Genotypes of female parent	$AA$	$Aa$	$aa$
Genotype frequency	$p^2$	$2pq$	$q^2$
Genotypes in S <sub>1</sub> families	$AA$	$\frac{1}{4}AA \frac{1}{2}Aa \frac{1}{4}aa$	$aa$
Frequency of A in S <sub>1</sub> family	1	$\frac{1}{2}$	0
S <sub>1</sub> family mean	$a$	$\frac{1}{2}d$	$-a$

$$\text{Mean of S}_1 \text{ families} = a(p - q) + dpq$$

$$\text{Variance of S}_1 \text{ families} = V_{G(S1)} = 2pqa_S^2 + (pqd)^2 = V_S + \frac{1}{4}V_D,$$

where  $\alpha_S = a + \frac{1}{2}d(q - p)$

Covariance of gene frequency and phenotypic value equals covariance of gene frequency and genotypic value =  $pq\alpha_S$ .

Regression ( $\beta$ ) of gene frequency on phenotypic value:

$\beta = pq\alpha_S/V_{P(S1)}$ , where  $V_{P(S1)}$  is the variance of phenotypic family means.

$V_{P(S1)} = V_{G(S1)} + V_E/r = V_S + \frac{1}{4}V_D + V_E/r$ , where  $V_E$  is the plot to plot environmental variation in the assessment trial and  $r$  is the number of replicates of the S<sub>1</sub> families in the trial.

### **Response to Selection**

The phenotypic difference between the mean of the selected group ( $M_S$ ) and the mean of the unselected population ( $M$ ) is the selection differential  $S$ , where the selected group can be individuals or families.

$$S = M_S - M$$

The standardized selection differential  $S/\sigma_P$  (in standard deviation units) is called the intensity of selection  $i$ .

The intensity of selection depends only on the proportion of the population selected and hence can be determined from tables of the properties of the normal distribution, provided of course the phenotypic values are normally distributed.

The expected change in allele 'A' frequency ( $\Delta p$ ) for selection differential  $S$  is simply determined from the slope of the linear regression line ( $\beta$ ), when we plot the graph of allele 'A' frequency against the genotypic values of the individuals or families assessed.

$$\Delta p = \beta S$$

The effect of changes in allele 'A' frequency ( $\Delta p$ ) on the population mean ( $\Delta M$  is the change in the population mean) can be derived using the differential calculus.

$$M = a(p - q) + 2dpq = a(2p - 1) + 2dp(1 - p), \text{ therefore}$$

$$dM/dp = 2a + 2d - 4dp = 2[a + d(q - p)] = 2\alpha, \text{ where } \alpha = a + d(q - p), \text{ and}$$

$$\Delta M = \Delta p \cdot dM/dp$$

The change in the population mean (the difference between the mean phenotypic value of the offspring of the selected parents and the whole of the parental generation before selection) is the response to selection ( $R$ ), hence

$$R = \Delta M = \Delta p \cdot dM/dp = \beta S \cdot dM/dp = \beta S \cdot 2\alpha = 2\beta\alpha S$$

Using the values of  $\beta$  from Box 10.1 we obtain the following formulae for the responses to selection, where  $\Sigma$  means summation over loci.

Mass selection :	$R = (\Sigma 2pq\alpha^2/V_P)S$	$= [V_A/(V_A + V_D + V'_E)]S$
Half-sib family selection :	$R = (\Sigma \frac{1}{2}pq\alpha^2/V_{P(HS)})S$	$= [\frac{1}{4}V_A/(\frac{1}{4}V_A + V_E/r)]S$
Full-sib family selection :	$R = (\Sigma pq\alpha^2/V_{P(FS)})S$	$= [\frac{1}{2}V_A/(\frac{1}{2}V_A + \frac{1}{4}V_D + V_E/r)]S$
$S_1$ family selection :	$R = (\Sigma 2pq\alpha_S\alpha/V_{P(S1)})S$	$= [V_{AS}/(V_S + \frac{1}{4}V_D + V_E/r)]S$

$V_A$  is the additive genetic variance,  $V_D$  is the dominance variance and  $V_E$  is the plot to plot environmental variation in the assessment trial where  $r$  is the number of replicates. For mass selection  $V'_E$  is the environmental component of the plant to plant variation; unless individuals are cloned and grown in rows or larger plots, when it can be replaced by  $V_E/r$ . For mass, half-sib family and full-sib family selection the numerator is  $S$  multiplied by  $V_A$ ,  $\frac{1}{4}V_A$  and  $\frac{1}{2}V_A$ , respectively. In other words, the response to selection is proportional to the additive genetic variance. The proportions in square brackets are referred to as the narrow sense heritabilities ( $h_n^2$ ) for individuals or families, as appropriate. The response to  $S_1$  family selection is more complicated because selfing changes the population mean so that the selection differential which changes the allele frequency is relative to this mean, but the response to selection is determined from one random mating population to the next. I have used the symbols  $V_{AS}$  and  $V_S$  to make this point (see Box 10.1). As mentioned earlier, the standardized selection differential ( $S/\sigma_P$ , where  $\sigma_P = V_P^{1/2}$ ) is called the intensity of selection  $i$ . Likewise one could divide the response to selection (in the unit of measurement of the trait) by the square root of the additive genetic variance ( $\sigma_A = V_A^{1/2}$ ) to obtain the standardized selection response ( $R/\sigma_A$ ); that is the selection response in standard deviations of the additive genetic variance as done by Wricke and Weber (1986) in their book. In this book, however, the unstandardized selection response is used.

## Theory Under Tetrasomic Inheritance

We will consider the simple model of a single locus with just two alleles and confine our attention to mass and half-sib family selection. We will also assume chromosomal inheritance, which occurs in alfalfa but not for example in potato. Alfalfa typically has only one chiasma per bivalent so that multivalents do not form, and hence there is no double reduction (Veronesi et al. 2010). It is assumed that during meiosis the four homologues of each chromosome, pair at random to give two bivalents.

With two alleles at a tetraploid locus there are five genotypes. Unfortunately different and potentially confusing notations have been used in the literature for their genetic values. I have chosen the one used by Wright (1979) but with '2a' and '-2a' used for the two homozygotes AAAA (quadruplex) and aaaa (nulliplex), respectively. I have not used the symbol  $d$  for the dominance deviation as the latter is not simply analogous to the diploid situation. There are three heterozygotes: AAa (triplex), Aaa (duplex) and Aaaa (simplex). They require three parameters and Wright (1979) used  $h$ ,  $v$  and  $w$  in the following combinations because they give results in a neat form: AAAa ( $a + 3h + v + w$ ), AAaa ( $4h$ ) and Aaaa ( $-a + 3h - v + w$ ).

The linear regression of genotypic value on the number of 'A' alleles in the genotype is derived in Box 10.2. The algebra is very tedious but does simplify to the given results.

### Box 10.2: Derivation of Regression ( $\beta$ ) of Allele Frequency on Phenotypic Value Under Tetrasomic Inheritance

Single locus with frequencies of alleles 'A' and 'a' equal to  $p$  and  $q$ , respectively ( $p + q = 1$ )

#### (a) Mass (MASS) selection

Genotypes of individuals	AAAA	AAa	Aaa	Aaaa	aaaa
Genotype frequency	$p^4$	$4p^3q$	$6p^2q^2$	$4pq^3$	$q^4$
Frequency of A	1	$\frac{3}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	0
Genotypic value	$2a$	$a + 3h + v + w$	$4h$	$-a + 3h - v + w$	$-2a$

$$\text{Mean frequency of 'A'} = p$$

$$\begin{aligned} \text{Mean of individuals} &= \text{population mean} \\ &= 2(p - q)a + 12pqrh + 4pq(p - q)v \\ &\quad + 4pq(1 - 2pq)w \end{aligned}$$

(continued)

**Box 10.2** (continued)

Covariance of allele 'A' frequency and phenotypic value equals covariance of allele 'A' frequency and genotypic value =

$$\begin{aligned}
 & p^4(1)(2a) + 4p^3q\left(\frac{3}{4}\right)(a + 3h + v + w) + 6p^2q^2\left(\frac{1}{2}\right)(4h) \\
 & + 4pq^3\left(\frac{1}{4}\right)(-a + 3h - v + w) + q^4(0)(-2a) - \\
 & p[2(p - q)a + 12pqh + 4pq(p - q)v + 4pq(1 - 2pq)w] \\
 & = pq[a + 3(q - p)h + (6pq - 1)v + (q - p)^3w] \\
 & = pq\alpha_t
 \end{aligned}$$

Regression ( $\beta$ ) of allele 'A' frequency on phenotypic value:

$\beta = pq\alpha_t/V_P$ , where  $V_P$  is the variance of phenotypic values.

$V_P = V_G + V'_E = V_A + V_D + V_T + V_Q + V'_E$ , where  $V_D$ ,  $V_T$  and  $V_Q$  are the components of variance for diallelic (dominance), triallelic and tetra-allelic interactions and  $V'_E$  is the environmental component of the plant to plant variation.

(b) Half-sib (HS) family selection

Genotypes of female parent	AAAA	AAa	Aaa	aaa	
Genotype frequency	$p^4$	$4p^3q$	$6p^2q^2$	$4pq^3$	
Genotypic value	$2a$	$a + 3h + v + w$	$4h$	$-a + 3h - v + w$	
				$-2a$	
<hr/>					
Pollen AA frequency $p^2$	AAAA	$\frac{1}{2}AAAA$	$(1/6)AAAA$	$\frac{1}{2}AAa$	
		$\frac{1}{2}AAa$	$(4/6)AAa$	$\frac{1}{2}Aaa$	
			$(1/6)AAaa$		
Pollen Aa frequency $2pq$	AAa	$\frac{1}{2}AAa$	$(1/6)AAa$	$\frac{1}{2}Aaa$	
		$\frac{1}{2}Aaa$	$(4/6)Aaa$	$\frac{1}{2}aaaa$	
			$(1/6)aaaa$		
Pollen aa frequency $q^2$	Aaa	$\frac{1}{2}AAaa$	$(1/6)AAaa$	$\frac{1}{2}aaaa$	
		$\frac{1}{2}aaaa$	$(4/6)aaaa$	$\frac{1}{2}aaaa$	
			$(1/6)aaaa$		
<hr/>					
Frequency of A in HS family	$\frac{1}{2}p + \frac{1}{2}$	$\frac{1}{2}p + \frac{3}{8}$	$\frac{1}{2}p + \frac{1}{4}$	$\frac{1}{2}p + \frac{1}{8}$	
HS family mean	$2pa + 2q(p + 2)h + 2pqv + 2pqw$	$(2p - \frac{1}{2})a + (1\frac{1}{2} + 2q(p + 1))h + (\frac{1}{2}(p - q) + pq)v + \frac{1}{2}w$	$(2p - 1)a + (2\frac{1}{2} + 2pq)h + (\frac{1}{2}(p - q)v + \frac{1}{2}(1 - pq)w$	$(2p - 1\frac{1}{2})a + (1\frac{1}{2} + 2p(1 + q))h + (\frac{1}{2}(p - q) - pq)v + \frac{1}{2}w$	$(2p - 2)a + 2p(2 + q)h - 2pqv + 2pqw$

(continued)

**Box 10.2** (continued)

Mean frequency of 'A' =  $p$

$$\begin{aligned}\text{Mean of HS families} &= \text{population mean} \\ &= 2(p - q)a + 12pqh + 4pq(p - q)v \\ &\quad + 4pq(1 - 2pq)w\end{aligned}$$

Covariance of gene frequency and phenotypic value equals covariance of gene frequency and genotypic value =  $\frac{1}{4}pq[a + 3(q - p)h + (6pq - 1)v + (q - p)^3w] = \frac{1}{4}pq\alpha_t$

Regression ( $\beta$ ) of gene frequency on phenotypic value:

$\beta = \frac{1}{4}pq\alpha_t/V_{P(HS)}$ , where  $V_{P(HS)}$  is the variance of phenotypic family means.

$V_{P(HS)} = V_{G(HS)} + V_E/r = \frac{1}{4}V_A + (1/36)V_D + V_E/r$ , where  $V_D$  is the component of variance for diallelic (dominance) interactions and  $V_E$  is the plot to plot environmental variation in the assessment trial and  $r$  is the number of replicates of the HS families in the trial.

The responses to mass and half-sib family selection are predicted as follows. The phenotypic difference between the mean of the selected group ( $M_S$ ) and the mean of the unselected population ( $M$ ) is the selection differential  $S$ , where the selected group is either individuals or half-sib families.

$$S = M_S - M$$

The standardized selection differential  $S/\sigma_P$  is the intensity of selection  $i$ .

The population is assumed to be in single locus equilibrium with genotype frequencies: AAAA ( $p^4$ ), AAaA ( $4p^3q$ ), AAaa ( $6p^2q^2$ ), Aaaa ( $4pq^3$ ) and aaaa ( $q^4$ ), where  $p$  and  $q$  are frequencies of 'A' and 'a'.

The expected change in allele 'A' frequency ( $\Delta p$ ) for selection differential  $S$  is simply determined from the slope of the linear regression line ( $\beta$ ), when we plot the graph of allele 'A' frequency against the genotypic values of the individuals or families assessed.

$$\Delta p = \beta S$$

The effect of changes in allele 'A' frequency ( $\Delta p$ ) on the population mean ( $\Delta M$  is the change in the population mean) can be derived using the differential calculus.

$$\begin{aligned}M &= 2(p - q)a + 12pqh + 4pq(p - q)v + 4pq(1 - 2pq)w \\ &= 2(2p - 1)a + 12p(1 - p)h + 4p(1 - p)(2p - 1)v \\ &\quad + 4p(1 - p)(1 - 2p + 2p^2)w, \text{ therefore}\end{aligned}$$

$$dM/dp = 4[a + 3(q - p)h + (6pq - 1)v + (q - p)^3w] = 4\alpha_t$$

where  $\alpha_t$  is the average effect of the allele substitution ‘ $a$ ’ to ‘ $A$ ’.

$$\Delta M = \Delta p \cdot dM/dp$$

The change in the population mean (the difference between the mean phenotypic value of the offspring of the selected parents and the whole of the parental generation before selection) is the response to selection ( $R$ ), hence

$$R = \Delta M = \Delta p \cdot dM/dp = \beta S \cdot dM/dp = \beta S \cdot 4\alpha_t = 4\beta\alpha_t S$$

Using the values of  $\beta$  from Box 10.2 we obtain the following formulae for the responses to selection, where  $\Sigma$  means summation over loci.

$$\text{Mass selection : } R = (\Sigma 4pq\alpha_t^2/V_P)S = (V_A/V_P)S$$

$$\text{Half-sib family selection : } R = (\Sigma pq\alpha_t^2/V_{P(HS)})S = (\frac{1}{4}V_A/V_{P(HS)})S$$

$V_A$  is the additive genetic variance,  $V_P$  is the phenotypic variance for individuals and  $V_{P(HS)}$  is the phenotypic variance for half-sib families. Wright (1979) showed that  $V_A = \Sigma 4pq\alpha_t^2$  and  $V_P = V_A + V_D + V_T + V_Q + V'_E$ , where  $V_D$ ,  $V_T$  and  $V_Q$  are the components of variance for diallelic (dominance), triallelic and tetra-allelic interactions and are functions of ( $h$ ,  $v$  and  $w$ ), ( $v$  and  $w$ ) and ( $w$ ), respectively, as well as of  $p$  and  $q$ . With just two alleles at a locus only one or two different alleles are involved in the triallelic and tetra-allelic interactions, but in general up to three and four different alleles can be involved.  $V'_E$  is the environmental component of the plant to plant variation.  $V_{P(HS)} = \frac{1}{4}V_A + (1/36)V_D + V_E/r$ , where  $V_E$  is the plot to plot environmental variation in the assessment trial and  $r$  is the number of replicates of the HS families in the trial (Wricke and Weber 1986).

We appear to have the same result for diploids and tetraploids, where  $S$  is multiplied by the heritability which is the additive genetic variance divided by the total phenotypic variance. The only apparent difference is that the additive genetic variance (and dominance variance) now contains a contribution from triallelic and tetra-allelic interactions and likewise the phenotypic variance. However, there is a complication. We assumed that after the selection of new parents, random mating produced a new population that was in single-locus equilibrium; true for diploids but not for tetraploids. In tetraploids, with chromosomal segregation, disequilibrium is reduced by one third with every generation of random mating, so three generations are required to reduce the disequilibrium to 3.7 % of its initial value (Crow and Kimura 1970). Hence we have predicted the equilibrium response to selection.

One simple way to determine the immediate response to mass selection is from the regression of offspring mean on midparental value. This differs in one respect

from our previous considerations where we practised selection and then random mated the selected individuals. We now assume that the individuals of the parent generation (mean  $M$ ) first mate at random to give full-sib families before we select on midparental values (mean of selected midparents  $M_S$ ). Although we are interested in the mean of the new population (i.e. all of the individuals in the new population, mean  $M_O$ ), we keep the family structure in order to determine the offspring-midparent covariance. The predicted mean of the new offspring population is:

$$M_O = M + \beta(M_S - M),$$

where

$M_O - M$  is the response to selection ( $R$ ),

$M_S - M$  is the selection differential ( $S$ ) and

$\beta$  is the regression coefficient, which is the offspring-midparent covariance divided by the phenotypic variance of midparental values which in turn is half the phenotypic variance of the parental values (individuals in unselected population).

The offspring-midparent covariance is  $(1/2)V_A + (1/6)V_D$  (Wright 1979).

The phenotypic variance of midparental values is  $(1/2)(V_A + V_D + V_T + V_Q + V'_E)$  (Kempthorne 1957). Hence:

$$R = \left[ \left( V_A + \frac{1}{3}V_D \right) / (V_A + V_D + V_T + V_Q + V'_E) \right] \cdot S$$

The immediate response to selection is greater than the equilibrium response achieved after a number of generations of random mating. Therefore the yield of an open-pollinated cultivar may fall slightly on multiplication to produce certified seed. The presence of  $1/3V_D$  in the numerator can be explained in terms of diallelic interactions being transmitted to the next generation through diploid gametes.

Likewise, the variance of half-sib family means (the covariance of half-sibs) is  $(1/4)V_A + (1/36)V_D$  (Kempthorne 1957). Nevertheless, the latter component is unlikely to seriously bias results and multiplying the estimated covariance of half-sibs by four [ $V_A + (1/9)V_D$ ] is still a good estimate of the additive genetic variance  $V_A$ . Furthermore, there will be little difference between the equilibrium response to half-sib family selection and the response predicted from half-sib family means.

Half-sib family selection:

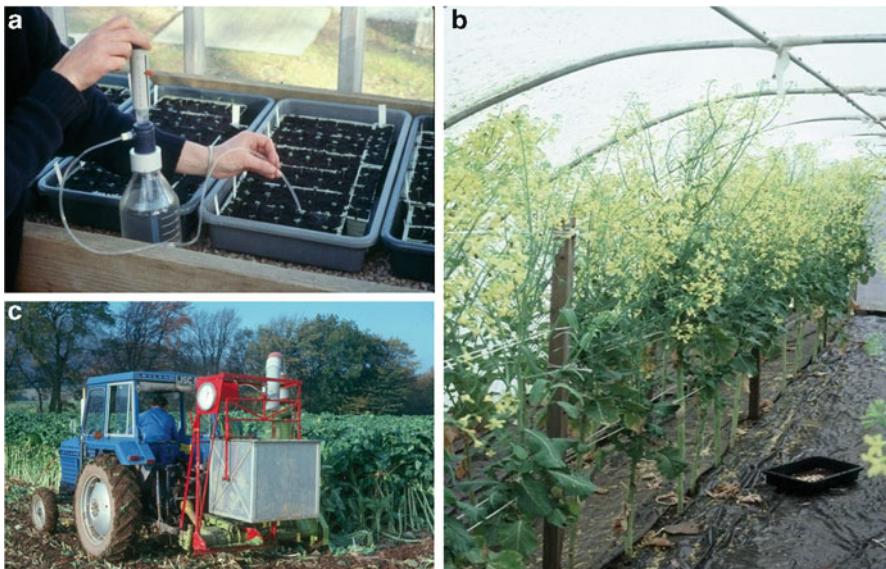
$$R = S \cdot [(1/4)V_A + (1/36)V_D] / [(1/4)V_A + (1/36)V_D + V_E/r]$$

where  $V_E$  is the plot to plot environmental variation in the assessment trial and  $r$  is the number of replicates of the HS families in the trial.

## Making Choices

Before starting a breeding programme it is sensible to do an experiment to estimate  $V_A$ ,  $V_D$  and  $V_{AS}$  so that responses to selection can be predicted and compared, and a sensible choice made.  $V_A$  and  $V_D$  can be estimated from the variances of half-sib and full-sib family means as explained in textbooks of Quantitative Genetics such as *The Genetical Analysis of Quantitative Traits* (Kearsey and Pooni 1996) and the fourth edition of *Introduction to Quantitative Genetics* (Falconer and Mackay 1996). The variances of these family means are also referred to as the covariances of half sibs and full sibs, respectively. The precision of estimates and the assumptions involved in extending the simple additive-dominance model to many loci are also discussed in these textbooks. What I have called  $V_{AS}$  is in fact twice the genetic covariance (i.e.  $2 \times \sum p q \alpha_S \alpha$ ) between either HS and S<sub>1</sub> family means or FS and S<sub>1</sub> family means (Bradshaw 1983). Once this is appreciated, practical methods of estimating the predicted response follow that involve selfing individual plants and also using them to produce half-sib and/or full-sib families.

With characters of high heritability, single plant selection may be best due to large population sizes and hence high selection intensities; for example, <1 % selected compared with 10 % in family selection. In contrast, with characters of low heritability, improvement should be faster with family selection than with single plant selection, because replicated trials allow the effects of environmental and genotype  $\times$  environmental interaction variation to be reduced. However, this does need to be quantified by experiment. Likewise, it is important to take into account practical considerations when comparing selection schemes, such as the time in years to complete one cycle and the ease of seed production. Hence one would not necessarily expect the breeders of maize, rye and kale to come to the same conclusions about the best scheme. With kale (Bradshaw and Wilson 2012) I found that full-sib family selection was better than half-sib family selection (Fig. 10.1) for all traits considered, despite the differing relative amounts of dominance variance and the variation in heritabilities. The intensity of selection was the same with both schemes because the extra work in producing full-sib families is not that onerous, and hence the same number of families can be produced and assessed in yield trials. Likewise in rye, many breeders changed from half-sib to full-sib family selection because of the greater expected response to selection, but producing the pair crosses does require considerably more work (Geiger and Miedaner 2009). The pair crosses have to be produced under pollination bags, but this can be done without emasculation as rye has an effective gametophytic incompatibility system. In rye, family selection involving yield assessment normally operates on a 4 year cycle, compared with two for kale, because of the need for an extra generation of seed production to produce enough for the yield trials. This is done by multiplication of the best families from observation plots through open-pollination in plots separated from each other by foliar isolation walls or cabins. The cycle can be shortened from 4 to 2 years if mother plant genotypes are quickly cloned in autumn between harvest and planting to produce about 20 plants



**Fig. 10.1** Kale population improvement. (a) Initial assessment for clubroot resistance of kale cultivars for foundation population. (b) Production of half-sib families in polythene tunnel using blowflies as pollinators. (c) Assessment the following year of half-sib families in replicated yield trial

per clone for seed production. Selection theory also helps the breeder make decisions about the optimum balance between the number of families to assess and their level of replication, and also the number of families to select. Kale examples are referred to in the paper by Bradshaw and Wilson (2012), but it is time to return to an economically more important crop, maize.

## Population Improvement in Practice

### *Maize*

More population improvement schemes have been done with maize than with any other crop, so it is worth looking at the results. Many open-pollinated cultivars were produced in the USA from 1865 to 1935 by simple mass selection, but average US grain yields did not improve (Troyer 2006). Hallauer and Carena (2009) attribute this to selection based on individual plants (low heritability), selection on plant and ear traits that did not affect grain yield (low correlations), inadequate control of male gametes (poor or no isolation) and poor plot techniques (poor control of local environmental effects). However, research from 1920 to 1950 concentrated on the development of double-cross hybrids and from 1935 to 1965 there were dramatic

increases in yield of over six tonnes per hectare. More sophisticated forms of population improvement were not examined in depth until there were concerns in the 1950s that yield increases might plateau. The context, however, was breeding better inbreds for hybrid cultivar production, rather than simple open-pollinated cultivars. Hence the population was often crossed with testers for assessment of combining ability, as explained in the Chap. 12. Nevertheless, the results are relevant and it is worth reading the review by Hallauer and Carena (2009) as well as *Quantitative Genetics in Maize Breeding* (Hallauer et al. 2010). Breeders will find reassurance that the methods work and guidance on the numbers of families to assess and select, levels of replication over sites and within sites, and likely progress both per cycle and per year. Breeders will not find the blueprint for the best population improvement scheme for all crops in all circumstances! Typically 100–500 families were assessed, often at three to six locations with two replications per location, and the best 10–30 families were selected to provide the parents for the next cycle. Progress was measured in one of two ways: either the same checks (controls) were included in each cycle of assessment, or the populations from each cycle were assessed in the same replicated trials. The latter often required a fresh seed production of each population to ensure that all populations had the same seed quality; and bulking within populations of equal quantities of seed from each plant to ensure that all were equally represented. The response to selection was usually averaged over cycles (or years) by linear regression analysis. Responses for grain yield were commonly around 3–5 % per cycle, and thus provided evidence of adequate additive genetic variance. Hallauer et al. (2010) reported a ratio of additive genetic variance to dominance variance of 1.6 for grain yield when averaged over many experiments, and higher values for traits other than yield. However, it is important to stress that breeders need to design their own programmes based on the reproductive biology of their crops and their resources, together with relevant genetic data, and also to monitor progress. If results prove disappointing a reassessment of the breeding programme will be required.

## ***Perennial Ryegrass***

Extensive population improvement followed by the production of open-pollinated and synthetic cultivars has been done in fodder crops and amenity grasses. Many examples can be found in Volume 5 of the Handbook of Plant Breeding (Boller et al. 2010a). Here we will briefly consider perennial ryegrass, alfalfa and clover.

About 50 % of the grass seed used in Europe is perennial ryegrass (*Lolium perenne*). Genetic gains in dry-matter yield over the past 50 years have been estimated at between 0.2 and 0.9 % per annum (Humphreys et al. 2010). These are only modest gains, but Humphreys et al. (2010) point out that changes in harvest index are limited in forages because all the above-ground biomass is harvested, and an increased shoot/root ratio is potentially undesirable due to possible loss of

efficiency in the uptake of water and plant nutrients. Average water-soluble carbohydrate concentration increased by 37 g/kg in cultivars bred between 1991 and 2000 whilst yield increased by 13 %.

Ravel and Charmet (1996) presented a comprehensive multisite recurrent selection strategy in perennial ryegrass along with a detailed analysis of the first 4-year-cycle of a polycross. Forty eight elite late flowering plants were selected from 19 out of 547 natural populations of perennial ryegrass collected in France during 1983 and 1984. Half-sib families were secured from a polycross mating design in 1988. Thirty three had enough seed for assessment as 2 replicates of 10 spaced plants at each of 8 locations from 1989, together with three replicates of sward plots at each of four of the locations. The spaced plants were assessed for seven traits (susceptibility to crown rust *Puccinia coronata*, early spring growth, aftermath (cutting) heading, summer growth, autumn growth, persistence and heading date), and the swards for dry-matter yields in spring and summer-autumn averaged over 5–9 cuts in both 1990 and 1991, as well as for persistency. Ten out of the 33 half-sib families were selected using an index of the three traits over the four locations, with trait × location combinations considered different traits and given equal economic weights. Within these families a total of 65 individual plants were selected from three locations (about 10 %) based on an individual selection index of up to six out of the seven traits. The subset of locations was chosen according to a multiplicative model decomposition of G × E interaction. Although a small experimental breeding programme, improvements of 10–15 % were expected for summer-autumn production and persistence, together with progress from the within family selection on crown rust tolerance, growth scores, leafiness and persistence. Discussions of population improvement methods in fodder crops commonly assume around 10,000 individual spaced plants for mass selection and around 100 progenies for family selection (Posselt 2010).

## **Alfalfa**

The following brief account is taken from the review by Veronesi et al. (2010). Alfalfa (*Medicago sativa*) is an autotetraploid ( $2n = 4x = 32$ ), perennial legume whose symbiont *Sinorhizobium meliloti* fixes nitrogen. It is cross-pollinated by bees and suffers severe inbreeding depression for fertility and vigour on self-pollination. Although domesticated several thousand years ago in Asia and spread worldwide, it was little-used until the mid 1700s when perennial forage legumes became widely used to produce hay for military horse herds and to improve soil fertility; alfalfa means horse fodder. Landrace populations developed from traditionally cultivated populations in local regions throughout the world. Use of molecular markers in recent years has revealed that most genetic variation in alfalfa resides within populations. Two interfertile subspecies, *sativa* and *falcata*, and their hybrids ( $\times$  *varia*) are recognized. Today alfalfa is the most important forage legume in a large part of the Americas and Europe. It is grown for hay, dehydrated forage, pellets, silage, and occasionally for grazing. It can be harvested for up to 4–5 years before

the stand deteriorates, although rotation to other crops after 2–3 years is common. There appears to have been only a small increase in dry-matter yields from breeding during the twentieth century, perhaps as little as 0.17–0.18 % per year in the USA. In contrast, progress has been made in selecting for improved winter survival and resistance to fungal diseases (such as anthracnose *Colletotrichum meliloti*), and in Europe to stem nematode (*Ditylenchus dipsaci*). The large majority of modern alfalfa cultivars are derived from phenotypic recurrent selection and synthetic strain building. As alfalfa is an autotetraploid, heterosis for yield should be maximized in a double-cross hybrid but they are difficult to produce and not yet economical. Interestingly, Veronesi et al. (2010) consider genomic selection an attractive proposition in alfalfa breeding by recurrent selection, particularly because extensive single nucleotide polymorphisms (SNPs) are likely to be available in the near future as DNA sequencing costs continue to fall.

### ***Red Clover***

The following brief account is taken from the review by Boller et al. (2010b). Red clover (*Trifolium pratense*) is a perennial forage legume of limited persistence, and hence mainly grown in grass-clover leys of 2–4 years duration for cutting to produce fresh or conserved fodder. It can be grown in pure stand but is best mixed with tall growing grasses. Its high protein content complements highly digestible grasses like Italian ryegrass. In symbiosis with *Rhizobium leguminosarum* biovar *trifolii*, it is capable of providing nearly 400 kg of fixed nitrogen per hectare per year. Red clover is a bee-pollinated, diploid ( $2n = 2x = 14$ ) outbreeding species with a strong gametophytic self-incompatibility system. It is thought to have originated from south-eastern Eurasia and has a cultivation history dating back to the thirteenth century; but came to prominence in Europe at the end of the sixteenth century when it replaced fallow land in the medieval three-course rotation. Locally adapted populations developed from the farming practice of re-sowing home-grown seed within the locality; and until about 1950, a great diversity of such landraces was cultivated in many regions of the world. These provided the germplasm for systematic breeding programmes that produced the high-quality cultivars which rapidly replaced landraces from the 1970s. Interestingly, substantial improvement in forage yield of red clover was obtained by inducing polyploidy, once reliable techniques became available in the 1950s using colchicine treatment of young seedlings. Disease resistance and persistence have also been improved but seed yield is markedly lower. Hence the relative market success of tetraploid cultivars ( $2n = 4x = 28$ ) depends on the willingness of farmers to pay a higher seed price. A second seed harvest year can increase total seed yield.

Recurrent mass and maternal line selection are the most frequently used breeding methods. At the tetraploid level, pairs of plants are manually crossed to produce a population of at least 50 crosses involving more than 30 unrelated individuals. The material then undergoes at least two generations of mass selection.

Selected individuals can be intermated by naturally occurring bee pollinators at isolated sites or in bee cages. Candidate cultivars are usually created by combining superior progenies of a number of elite individuals (parents), either by open-pollination or pair crosses. The progenies are tested in small plots or rows and the best ones intermated to form the new candidate cultivar. Hence the resulting synthetic cultivar has a fixed number of half-sib or full-sib families as components, rather than parental clones, because vegetative propagation is difficult and long-term maintenance of clones is risky.

### **White Clover**

The following brief account is taken from the review by Abberton and Marshall (2010). White clover (*Trifolium repens*) is an allotetraploid ( $2n = 4x = 32$ ) species that displays disomic inheritance, its likely diploid progenitors being *T. pallescens* and *T. occidentale*. It is a bee-pollinated, outbreeding species with a single-locus gametophytic self-incompatibility system. White clover spreads by means of stolons (horizontal stems) and thus has many active growing points which are the key to its persistence and tolerance of defoliation. It is grown in pastures in mixed swards, most commonly with perennial ryegrass, which are grazed by sheep or cattle. Leaf size is closely related to the size of stolons and determines suitability for a particular livestock system; for example, small-leaf cultivars are considered best for continuous hard grazing by sheep. The overall breeding objective is therefore a cultivar that gives a balanced sward for 10 or more years, with a reliable white clover contribution (for example, 30 %) under grazing management. An important part of this system is the impact of the atmospheric nitrogen fixation by the clover and associated Rhizobia bacteria on the nitrogen cycle, including nitrogen transfer from clover to companion grass and any subsequent crop. Winter hardiness (or summer survival in some countries) and resistance to pests such as slugs, weevils and nematodes are also important breeding objectives. Ultimately the farmer is seeking improved meat and milk production and quality. The cyanogenesis potential of white clover is given consideration in some breeding programmes because of concerns about its effects on large herbivores. The genetics of this trait was dealt with in Chap. 5.

The main breeding method in white clover has been population improvement followed by synthetic cultivar production from a small number of vegetatively propagated mother plants. New Zealand and Denmark are the largest producers of certified seed, accounting for over 80 % of world production. Single spaced plants in rows can be assessed for leaf size and other general characteristics. However, for agronomic and performance traits, evaluations are done in swards with the companion grass over a period of at least 3 years, with an appropriate management system involving sheep or cattle grazing. A rare dominant gene ( $S_f$ ) conferring self-compatibility has been discovered and inbred lines have been produced. Better-parent heterosis has been found for dry-matter yield; but hybrid cultivars have not been produced to date (Michaelson-Yeates et al. 1997).

## Mixed Selfing and Random Mating

Not all species are strictly cross (or self)-fertilizing. Some, as we have seen, have a breeding system which involves a mixture of selfing and crossing. Today plant breeders tend to impose strict cross- or self-fertilization to produce high yielding hybrid, synthetic, or inbred line cultivars. They do this by emasculation to prevent self-pollination or bagging flowers to prevent cross-pollination. Furthermore, when recessive genic male sterility (*msms*) is available in an inbreeding crop, it can be used to create an outbreeding population by only harvesting seed from male sterile plants, which will comprise half of the population after three generations (Suneson 1951). The initial hybrids between male fertile lines and the male sterile one will all be *MsMs*. On selfing, one quarter of the plants will be male sterile (*msms*). They will be pollinated on average by pollen from one *MsMs* and two *Msms* plants so that one third of their offspring will be male sterile. These will be pollinated by *Msms* plants so that half of their offspring will be male sterile, as will subsequent generations. These single plant progenies are half-sib families and hence the population can be improved by half-sib family selection. Inbred lines can subsequently be produced for evaluation by selfing the offspring of true-breeding male fertile plants (*Msms* → *MsMs*).

In the past population improvement schemes using natural pollination were used to produce open-pollinated cultivars, and such schemes could be used today for further germplasm enhancement. Wright and Cockerham (1986) therefore extended the theory of half- and full-sib (outcrossed) family selection to accommodate partial selfing in naturally-pollinated maternal families and pair crossed ones without exclusion of selfing. Previously they (Wright and Cockerham 1985) had developed the theory of mass selection with partial selfing. They considered permanent as well as immediate responses. However, predictions and hence comparisons of responses to selection require estimation of covariances of relatives in terms of six quadratic components for a model with arbitrary additive and dominance effects, number of alleles and non-epistatic loci. Estimation of these components requires extensive experimentation and the estimates have large standard errors (Cockerham and Matzinger 1985). Hence practical plant breeders might prefer to be guided by computer simulations of possible schemes for their crops for a range of genetic models. Such schemes should include ones with selfed progenies as theory does predict their efficiency for the improvement of outcrossed populations (Wright and Cockerham 1986).

## Linkage Disequilibrium

The formulae we have derived for predicting responses to selection typically provide a reasonable description of the first few generations, but strictly speaking apply to just one generation of selection. This is because the effect of selection is to change allele frequencies and hence the genetic variance and heritability. Furthermore, with many loci, selection generates linkage disequilibrium which reduces the

additive genetic variance and hence heritability and the rate of progress. This phenomenon is known as the Bulmer effect and the theory can be found in his book *The Mathematical Theory of Quantitative Genetics* (Bulmer 1980). Selection can also drive the parent-offspring regression from linearity (Walsh 2004).

The theory is quite complicated so here we will just demonstrate the Bulmer effect with two simple examples given in Box 10.3. In the example with mass selection, we see the allele frequencies of 'A' and 'B' increase from 0.5 to 0.8 and the population mean increase from 0 to  $1.2a$ . The population variance is expected to decrease from  $a^2$  to  $0.64a^2$  as a result of the change in allele frequencies, but is in fact  $0.48a^2$  as a result of a further reduction (25 %) from the linkage disequilibrium generated. It can also be seen in Box 10.3 that the disequilibrium ( $D$ ) is halved as a result of recombination when the next set of gametes is produced, as expected for unlinked loci. However, further selection can generate more disequilibrium. Eventually a balance (equilibrium) is reached between the increase in disequilibrium due to selection and the decrease due to recombination. Most of the increase occurs in the first generation and with unlinked loci the balance is almost achieved after three or four generations. Hence if changes in allele frequency are very small (unlike in our artificial example), the response to selection can be fairly constant for a number of generations, until changes in allele frequency are large enough to take effect. We shall see this at the end of this chapter when we consider a long term selection experiment in maize. Walsh (2004) gives a theoretical example of where the equilibrium response is 85 % of the initial response. The example of half-sib family selection makes the additional point that the between family variance, but not the within family variance, is affected by linkage disequilibrium. The change in gene frequencies has increased the population mean from 0 to  $0.6a$  and reduced the between family variance from  $0.25a^2$  to  $0.2275a^2$ , but the Bulmer effect has reduced it further (by about 19 %) to  $0.185a^2$ . In contrast, the change in gene frequencies has reduced the within family variance from  $0.75a^2$  to  $0.6825a^2$  but there is no further reduction from the Bulmer effect (values would be identical in an infinite population with infinite number of genes). The Bulmer effect is smaller when heritability and selection intensity are lower.

### **Box 10.3: The Bulmer Effect with Mass Selection and Half-Sib Family Selection**

Let us start with a population in equilibrium and consider two unlinked loci, each with two alleles at equal frequencies ( $A = a = \frac{1}{2}$ ) and ( $B = b = \frac{1}{2}$ ), and a simple additive genetic model with equal effects at the two loci ( $AA = BB = a$ ;  $Aa = Bb = 0$  and  $aa = bb = -a$ ).

Hence population mean  $M_0 = 0$  and population variance  $V_0 = a^2$  (sum over two loci).

(a) Mass selection

Let us select genotypes with values  $\geq a$ :  $AABB$ ,  $AABb$  and  $AaBB$ , and mate them at random:

(continued)

**Box 10.3** (continued)

Genotype	<i>AABB</i>	<i>AABb</i>	<i>AaBB</i>			
Genotypic value	$2a$	$a$	$a$			
Frequency $\times 5$	1	2	2			
Gametes	$AB$	$\frac{1}{2}AB \frac{1}{2}Ab$	$\frac{1}{2}AB \frac{1}{2}aB$			
Random mate	$(3AB \ 1Ab \ 1aB) \times (3AB \ 1Ab \ 1aB)/25$					
Offspring						
Genotype	$AB/AB$	$AB/Ab$	$AB/aB$	$Ab/Ab$	$AB/ab$	$aB/aB$
Genotypic value	$2a$	$a$	$a$	0	0	0
Frequency	9/25	6/25	6/25	1/25	2/25	1/25

Frequency of 'A' = frequency of 'B' = 0.8

Population mean  $M_1 = 1.2a$  and population variance  $V_1 = 0.48a^2$

Variance of equilibrium population  $V_e = 0.64a^2$

Bulmer effect is  $0.16a^2$  or 25 % reduction.

Linkage disequilibrium ( $D$ )

$$D = P(AB)P(ab) - P(AB)P(ab) :$$

Gametes	$AB$	$Ab$	$aB$	$ab$	
From parents	3/5	1/5	1/5	0/5	$D = -0.04$
From offspring	31/50	9/50	9/50	1/50	$D = -0.02$

(b) Half-sib family selection

Initial population:

Genotype	<i>AABB</i>	<i>AABb</i>	<i>AAAb</i>	<i>AaBB</i>	<i>AaBb</i>	<i>Aabb</i>	<i>aaBB</i>	<i>aaBb</i>	<i>aabb</i>
Genotypic value	$2a$	$a$	0	$a$	0	$-a$	0	$-a$	$-2a$
Frequency $\times 16$	1	2	1	2	4	2	1	2	1

Half-sib families produced from population:

Female		<i>AABB</i>	<i>AABb</i>	<i>AAAb</i>	<i>AaBB</i>	<i>AaBb</i>	<i>Aabb</i>	<i>aaBB</i>	<i>aaBb</i>	<i>aabb</i>
Pollen	$\frac{1}{4}AB$	x	xx		xx					
	$\frac{1}{4}Ab$	x	xx		xx					
	$\frac{1}{4}aB$	x	xx		xx					
	$\frac{1}{4}ab$	x	xx		xx					
Family mean		$a$	$\frac{1}{2}a$	0	$\frac{1}{2}a$	0	$-\frac{1}{2}a$	0	$-\frac{1}{2}a$	$-a$

Mean of half-sib families = 0; Variance between half-sib families =  $\frac{1}{4}a^2$ ; Variance (average) within half-sib families =  $\frac{3}{4}a^2$ .

(continued)

**Box 10.3** (continued)

Let us select families with means  $\geq \frac{1}{2}a$ : female parental genotypes  $AABB$ ,  $AABb$  and  $AaBB$ .

( $x$  above marks number of genotypes in each family, and as frequency of  $AABB$  is half that of other two female parents, there are 20 genotypes ( $x$ 's) of equal frequency).

Frequencies of selected genotypes will be:

	$AABB$								
Genotypic value	$2a$	$a$	0	$a$	0	$-a$	0	$-a$	$-2a$
Frequency $\times 20$	3	4	1	4	5	1	1	1	0

Frequency of ' $A$ ' = frequency of ' $B$ ' = 0.65 (increase from 0.50)

Frequency of ' $a$ ' = frequency of ' $b$ ' = 0.35 (decrease from 0.50)

Mean of selected individuals =  $0.6a$  (increase from 0)

Half-sib families are now produced from these selected individuals:

Female		$AABB$	$AABb$	$AAbb$	$AaBB$	$AaBb$	$Aabb$	$aaBB$	$aaBb$	$aabb$
Frequency $\times 20$		3	4	1	4	5	1	1	1	0
Pollen $\times 20$	$8\frac{1}{4}AB$									
	$4\frac{3}{4}Ab$									
	$4\frac{3}{4}aB$									
	$2\frac{1}{4}ab$									
Family mean		$1.3a$	$0.8a$	$0.3a$	$0.8a$	$0.3a$	$-0.2a$	$0.3a$	$-0.2a$	-

Mean of half-sib families = Mean of selected individuals =  $0.6a$  (increase from 0)

Variance between half-sib families =  $0.185a^2$  Expected in equilibrium =  $0.2275a^2$

Variance (average) within half-sib families =  $0.685a^2$  Expected in equilibrium =  $0.6825a^2$

Total variance between individuals =  $0.87a^2$  Expected in equilibrium =  $0.91a^2$

Mackay and Gibson (1993) have shown that failure to take into account the effects of gametic-phase disequilibrium can result in substantial overestimation of the response to recurrent selection in plants which could affect decisions on the merits of alternative breeding schemes; their particular interest being sugar beet. In contrast, for a given scheme, ignoring gametic-phase disequilibrium had only small

effects on optimum resource allocation (e.g. optimum numbers of families and replicates for given number of plots). Once again, I think that it is reasonable to conclude from the results of Mackay and Gibson (1993) that practical plant breeders might like to be guided by computer simulations of possible schemes for their crops for a range of genetic models.

## Epistasis

We will use mass selection to consider the complication of epistasis. We now have to determine the expected changes in allele frequencies at two loci  $A$  and  $B$ , namely  $\Delta p$  (for allele 'A' at locus  $A$ ) and  $\Delta u$  (for allele 'B' at locus  $B$ ), from the slopes of the linear regression lines ( $\beta_A$  and  $\beta_B$ ), when we plot the graphs of allele 'A' and allele 'B' frequencies against the genotypic values of the individuals assessed.

$$\Delta p = \beta_A S \text{ and } \Delta u = \beta_B S$$

The effect of changes in both allele 'A' frequency ( $\Delta p$ ) and allele 'B' frequency ( $\Delta u$ ) on the population mean ( $\Delta M$  is the change in the population mean) can be derived using the differential calculus.

$$\Delta M = \Delta p.dM/dp + \Delta u.dM/du$$

The population mean for two loci in linkage equilibrium that display epistasis is shown in Box 10.4 along with  $\beta_A$  and  $\beta_B$  for mass selection.

$$dM/dp = 2\alpha_A \text{ and } dM/du = 2\alpha_B, \text{ where } \alpha_A \text{ and } \alpha_B \text{ are also given in Box 10.4.}$$

The change in the population mean (the difference between the mean phenotypic value of the offspring of the selected parents and the whole of the parental generation before selection) is the response to selection ( $R$ ). Hence:

$$R = \Delta M = \Delta p.dM/dp + \Delta u.dM/du = \beta_A S.2\alpha_A + \beta_B S.2\alpha_B = 2\beta_A \alpha_A S + 2\beta_B \alpha_B S$$

Using the values of  $\beta$  from Box 10.4 we obtain the following formulae for the responses to mass selection, where  $\Sigma$  means summation over loci.

$$\begin{aligned} R &= (\Sigma 2pq\alpha_A^2/V_P)S + (\Sigma 2uv\alpha_B^2/V_P)S \\ &= [V_A/(V_A + V_D + V_{AA} + V_{AD} + V_{DD} + V'_E)].S \end{aligned}$$

where  $V_A$  is the additive genetic variance,  $V_D$  is the dominance variance and  $V_{AA}$ ,  $V_{AD}$  ( $V_{AD}$  includes  $V_{DA}$ ) and  $V_{DD}$  are additive  $\times$  additive, additive  $\times$  dominance and dominance  $\times$  dominance interaction variances.

**Box 10.4: Mass Selection: Population with Two Loci in Linkage Equilibrium and Displaying Epistasis (Frequencies of Alleles  $A$ ,  $a$ ,  $B$  and  $b$  are  $p$ ,  $q$ ,  $u$  and  $v$ )**

Genotypes	Frequencies	Frequency of $A$	Frequency of $B$	Genotypic value
$AABB$	$p^2u^2$	1	1	$a_A + ab + aa_{AB}$
$AAbB$	$2p^2uv$	1	$\frac{1}{2}$	$a_A + d_B + ad_{AB}$
$AAbb$	$p^2v^2$	1	0	$a_A - a_B - aa_{AB}$
$AaBB$	$2pqu^2$	$\frac{1}{2}$	1	$a_B + d_A + ad_{BA}$
$AaBb$	$4pquv$	$\frac{1}{2}$	$\frac{1}{2}$	$d_A + db + dd_{AB}$
$Aabb$	$2pqv^2$	$\frac{1}{2}$	0	$-a_B + d_A - ad_{BA}$
$aaBB$	$q^2u^2$	0	1	$-a_A + a_B - aa_{AB}$
$aaBb$	$2q^2uv$	0	$\frac{1}{2}$	$-a_A + d_B - ad_{AB}$
$aabb$	$q^2v^2$	0	0	$-a_A - a_B + aa_{AB}$

Mean of individuals = population mean =  $M$

$$M = (p - q)a_A + 2pqd_A + (u - v)a_B + 2uvd_B + (p - q)(u - v)aa_{AB} + 2pq(u - v)ad_{BA} + 2uv(p - q)ad_{AB} + 4pquvdd_{AB}.$$

Variance of individuals =  $V_G = V_A + V_D + V_{AA} + V_{AD} + V_{DD}$

where  $V_A$  is the additive genetic variance,  $V_D$  is the dominance variance and  $V_{AA}$ ,  $V_{AD}$  ( $V_{AD}$  includes  $V_{DA}$ ) and  $V_{DD}$  are additive  $\times$  additive, additive  $\times$  dominance and dominance  $\times$  dominance interaction variances.

$$V_A = 2pq\alpha_A^2 + 2uva_B^2, \text{ where}$$

$$\alpha_A = a_A + (u - v)aa_{AB} + 2uvad_{AB} + (q - p)(d_A + (u - v)ad_{BA} + 2uvdd_{AB}) \text{ and}$$

$$\alpha_B = a_B + (p - q)aa_{AB} + 2pqad_{BA} + (v - u)(d_B + (p - q)ad_{AB} + 2pqdd_{AB})$$

Covariance of allele 'A' frequency and phenotypic value equals covariance of allele 'A' frequency and genotypic value =  $pq\alpha_A$

Covariance of allele 'B' frequency and phenotypic value equals covariance of allele 'B' frequency and genotypic value =  $uv\alpha_B$

Regressions ( $\beta_A$  and  $\beta_B$ ) of allele 'A' frequency and allele 'B' frequency on phenotypic value:

$\beta_A = pq\alpha_A/V_P$  and  $\beta_B = uv\alpha_B/V_P$ , where  $V_P$  is the variance of phenotypic values.

$$V_P = V_G + V'_E = V_A + V_D + V_{AA} + V_{AD} + V_{DD} + V'_E$$

Hence we appear to have the same result as before, where the heritability is the additive genetic variance divided by the total phenotypic variance, although the additive genetic variance now contains a contribution from non-allelic interactions and the phenotypic variance includes components due to interactions between pairs of loci. However, there is a complication. We determined the change in means for two populations in linkage equilibrium, but selection creates disequilibrium which is not dissipated by one round of random mating. Hence the immediate response to selection is different to the equilibrium response, as we found with autotetraploids.

### ***Immediate Versus Equilibrium Response to Selection***

Once again the immediate response to selection can be determined from the regression of offspring mean on midparental value. We again assume that the individuals of the parent generation (mean  $M$ ) first mate at random to give full-sib families before we select on midparental values (mean of selected midparents  $M_S$ ). Although we are interested in the mean of the new population (i.e. all of the individuals in the new population, mean  $M_O$ ), we keep the family structure in order to determine the offspring-midparent covariance. The predicted mean of the new offspring population is:

$$M_O = M + \beta(M_S - M),$$

where

$M_O - M$  is the response to selection ( $R$ ),  
 $M_S - M$  is the selection differential ( $S$ ) and

$\beta$  is the regression coefficient, which is the offspring-midparent covariance divided by the phenotypic variance of midparental values which in turn is half the phenotypic variance of the parental values (individuals in unselected population).

The offspring-midparent covariance is  $\frac{1}{2}V_A + \frac{1}{4}V_{AA}$  (Crow and Kimura 1970).

The phenotypic variance of midparental values is  $\frac{1}{2}(V_A + V_D + V_{AA} + V_{AD} + V_{DD} + V'_E)$ . Hence:

$$R = [(V_A + \frac{1}{2}V_{AA}) / (V_A + V_D + V_{AA} + V_{AD} + V_{DD} + V'_E)].S$$

The difference between the immediate and equilibrium responses is therefore  $\frac{1}{2}V_{AA}$  in the numerator, where  $V_{AA} = 4pquv[aa_{AB} + (q - p)ad_{BA} + (v - u)ad_{AB} + (q - p)(v - u)dd_{AB}]^2$  (Wright 1979).

The main practical implication is that a breeder might see a small fall in the yield of an open-pollinated cultivar when it is multiplied for a number of generations.

### ***The Importance of Epistasis in Practice***

In their book *An Introduction to Population Genetics Theory*, Crow and Kimura (1970) make a number of important points about epistasis. The method of least squares for partitioning the genetic variation in a population takes up as much variance as possible in the additive term, then as much of the remainder as possible with the dominance deviation, and what is left over is attributed to epistasis. If one has classical complementary epistasis (9:7 ratio) and classical duplicate epistasis (15:1 ratio) with allele frequencies ( $p_A = q_B = 0.459$  and  $p_A = q_B = 0.159$ , respectively) adjusted so that the two phenotypes (scores 1 and -1) are equally frequent in both examples, and the measurements are scaled to give a total variance of 1, despite complete epistasis, only 17 % of the total variance appears in the epistatic term. The additive genetic variances make up 58 % and 76 %, and the dominance variances 25 % and 7 %, respectively. Furthermore, Crow and Kimura (1970) give an example of where ‘ignoring’ epistasis doesn’t cause a very large error in heritability measurements or predictions based on them. In other words, for practical purposes, good estimates of twice the offspring-midparent covariance ( $V_A + \frac{1}{2}V_{AA}$ ), or four times the covariance between half-sibs ( $V_A + \frac{1}{4}V_{AA}$ ), are adequate estimates of the additive genetic variance ( $V_A$ ). As there are also practical difficulties in measuring epistasis, breeders tend to ‘ignore’ epistasis in predicting the response to selection. For example, Hallauer and Carena (2009) comment that the more complex mating designs needed to estimate the epistatic components of variance have not been successful in maize.

## **Selection for More Than One Trait**

Up until now in this chapter we have talked about assessment trials, phenotypic differences and changes in population means, without specifying the phenotypes being selected. In a practical breeding programme the breeder will most likely want to improve yield, quality traits and resistances to the most serious abiotic and biotic stresses. But how does the breeder combine information on more than one trait when deciding which individuals or families to select as parents of the next generation? Three strategies can be used: tandem selection, independent culling levels and index selection (Wricke and Weber 1986). In tandem selection, only one character is selected in each cycle, say trait one in the first cycle, trait two in the second, and so on; but this is not an efficient procedure. With independent culling levels, all genotypes are discarded which have a phenotypic value below the culling level for at least one trait; for example, low yield, poor quality or poor disease resistance. This is the method that is most commonly used in plant breeding to select potential cultivars, often for different combinations of traits at different stages in a breeding cycle (for example, over the selfing and clonal generations with inbreeding and vegetatively propagated crops, respectively). We will look at

independent culling levels in this context in the next chapter. The theory for optimizing the culling levels for population improvement is difficult and not considered further in this book. We will now focus on the potentially most efficient method, use of a selection index.

## ***Selection Indices***

The aim of index selection is to improve several traits simultaneously in such a way as to make the biggest possible improvement in overall merit. For example, in kale for feeding to cattle and sheep, one wants to increase digestible organic matter yield whilst lowering the contents of known antimetabolites (S-methyl cysteine sulphoxide, the haemolytic factor, and indolyl glucosinolates, source of nitriles and goitrogenic thiocyanate ion) in order to improve the economics of milk and meat production (Barry 2013). Ideally one would weight the genetic improvement of each trait according to its effect on this economic value, but in the absence of such information, the weighting will be subjective, although based on the latest information about feeding value for ruminant livestock (Barry 2013). The basic theory and use of various indices can be found in the review by Lin (1978) and in books such as those by Wricke and Weber (1986) and Falconer and Mackay (1996). We will concentrate on the use of the Optimum Index of Smith (1936) for three traits in half-sib family selection of kale (Bradshaw 1987). Other indices were explored and discussed by Bradshaw (1987) who concluded that there was no good reason to reject the use the optimum index, despite concerns over the lack of objective economic values and the accuracy of the index.

### ***Optimum Index of Smith for Half-Sib Family Selection in Kale***

The breeding value for merit is symbolized by  $H$  (genotypic index) and for half-sib families this aggregate breeding value is the sum of the true family means of all the traits ( $y$ 's) to be improved, as deviations from the population mean, weighted ( $a$ 's) by their relative economic values. Hence for three traits:

$$H = a_1 y_1 + a_2 y_2 + a_3 y_3$$

$$I = b_1 x_1 + b_2 x_2 + b_3 x_3$$

The phenotypic index ( $I$ ) for the improvement of merit is the sum of the phenotypic family means ( $x$ 's), as deviations from the population mean, weighted ( $b$ 's) in such a way as to maximize the correlation between  $I$  and  $H$ , so that selection using  $I$  maximizes the change in  $H$ . This is equivalent to minimizing the sum of squared deviations in the regression of  $H$  on  $I$  (i.e.  $\Sigma(I - H)^2$ ). This is done by a standard statistical procedure in which the partial derivatives with respect to the  $b$ 's are set

equal to zero. The result is a set of simultaneous equations, as many as there are traits, whose solution gives the values of the  $b$ 's to be used in the index. The  $b$ 's are the partial regression coefficients of the individual families' breeding values on each measurement. In matrix algebra:

$$\mathbf{b} = \mathbf{P}^{-1}\mathbf{G}\mathbf{a}$$

where  $\mathbf{P}$  and  $\mathbf{G}$  are the phenotypic and additive genetic variance-covariance matrices. With no phenotypic or genotypic covariances (correlations) between traits this simplifies to:

$$\begin{array}{rcl} |b_1| & = & |1/P_1 \quad 0 \quad 0| \quad |G_1 \quad 0 \quad 0| \quad |a_1| & = & |a_1 G_1 / P_1| \\ |b_2| & = & |0 \quad 1/P_2 \quad 0| \quad |0 \quad G_2 \quad 0| \quad |a_2| & = & |a_2 G_2 / P_2| \\ |b_3| & = & |0 \quad 0 \quad 1/P_3| \quad |0 \quad 0 \quad G_3| \quad |a_3| & = & |a_3 G_3 / P_3| \end{array}$$

In other words, the weight for each trait is its economic weight multiplied by its heritability, which makes sense. When both desirable and undesirable correlations occur between traits, these are accommodated by the index. Furthermore, with half-sib families a new index can be calculated every cycle.

The predicted response to index selection ( $R$ ) is estimated from the regression of  $H$  on  $I$  multiplied by the selection differential  $\Delta I$ , i.e. the mean of families selected minus the mean of all families. The regression coefficient is the covariance of  $H$  and  $I$  divided by the variance of  $I$  ( $\sigma_I^2$ ). The covariance can be shown to be  $\mathbf{a}'\mathbf{G}\mathbf{b}$ . Hence:

$$R = \mathbf{a}'\mathbf{G}\mathbf{b} \cdot \Delta I / \sigma_I^2$$

The responses to selection for individual traits ( $R_1$ ,  $R_2$  and  $R_3$ ) are:

$$\mathbf{R} = \mathbf{G}\mathbf{b} \cdot \Delta I / \sigma_I^2, \text{ where } \mathbf{R} \text{ is a column vector with elements } R_1, R_2 \text{ and } R_3.$$

$$R = a_1 R_1 + a_2 R_2 + a_3 R_3$$

Calculation of the optimum index for one cycle of half-sib family selection in kale is shown in Box 10.5. In the absence of objective economic weights, equal weights in terms of phenotypic standard deviations were used ( $1/\sigma_P$ ), although it could be argued that these should be multiplied by the number of standard deviations required to equal control cultivars or to exceed them by a desired amount. The more traits that one includes in the index, the less progress one will make for each trait in each cycle. Hence the predicted responses are very useful in helping a breeder to get a feel for what can realistically be achieved in a given period of time, and may even lead to a re-appraisal of breeding objectives. For example, including fewer traits in the index but monitoring correlated responses in other traits. Certainly in kale it seemed best to focus on around six important traits, and this was borne out in practice (Bradshaw and Wilson 2012). The other useful conclusion that came from further analysis of the trial in Box 10.5 concerned the allocation of resources. For a fixed number of plots in the assessment trial, it was found

undesirable to increase the number of replicates beyond the minimum of two required to estimate the additive genetic variance-covariance matrix. This is because fewer families would be assessed and this reduces the overall and the individual responses to selection with traits of moderate heritability.

**Box 10.5: Calculation of the Optimum Index for One Cycle of Half-Sib Family Selection in Kale**

In 1985, near Dundee (Scotland), 84 half-sib families of kale were grown in a randomized complete block trial with two replicates (Bradshaw 1987). All 168 plots were assessed for thiocyanate ion ( $\text{SCN}^-$ ) content, S-methyl cysteine sulphoxide (SMCO) content and digestible organic-matter yield (DOMDY).

The phenotypic (**P**) and additive genetic (**G**) variance-covariance matrices for the three traits were estimated from analyses of variance and covariance for single traits and pairs of traits, respectively.

The relative economic values (*a*) chosen were equal in phenotypic standard deviation units ( $\sigma_P$ ):  $a = 1/\sigma_P$

The selection score (*I*) for each family was:

$I = b_1x_1 + b_2x_2 + b_3x_3$ , where traits 1–3 were  $\text{SCN}^-$ , SMCO and DOMDY, and  $x$  was the phenotypic deviation of the half-sib family mean (mean of 2 replicates) from the population mean.

The weights (*b*) were calculated as follows:

$$\mathbf{b} = \mathbf{P}^{-1}\mathbf{G}\mathbf{a}$$

$$\begin{aligned} \mathbf{P} = & \begin{array}{ccc} 82.887604 & 0.391451 & -0.735049 \\ 0.391451 & 0.015766 & -0.002438 \\ -0.735049 & -0.002438 & 0.283033 \end{array} \end{aligned}$$

$$\begin{aligned} \mathbf{G} = & \begin{array}{ccc} 52.529266 & 0.033767 & -0.959458 \\ 0.033767 & 0.006311 & -0.010993 \\ -0.959458 & -0.010993 & 0.143162 \end{array} \end{aligned}$$

$$\begin{aligned} \mathbf{P}^{-1} = & \begin{array}{ccc} 0.013975 & -0.341816 & 0.033348 \\ -0.341816 & 71.873383 & -0.268547 \\ 0.033348 & -0.268547 & 3.617447 \end{array} \end{aligned}$$

$$\begin{aligned} \mathbf{P}^{-1}\mathbf{G} = & \begin{array}{ccc} 0.690534 & -0.002052 & -0.004876 \\ -15.270736 & 0.444977 & -0.500608 \\ -1.728119 & -0.040336 & 0.488836 \end{array} \end{aligned}$$

(continued)

**Box 10.5** (continued)

$\mathbf{a} = -0.10984$	$\mathbf{P}^{-1}\mathbf{G}\mathbf{a} = -0.068673$	<b>Heritability</b>	$\text{SCN}^-$	0.63
- 7.96415	- 2.807499		$\text{SMCO}$	0.40
1.87967	1.429922		$\text{DOMDY}$	0.51

$$I = -0.0687(\text{SCN}^- - 50.86) - 2.8075(\text{SMCO} - 0.665) \\ + 1.4299(\text{DOMDY} - 9.59)$$

Predicted responses to selection of best 14 out of the 84 families were calculated as follows:

$$\mathbf{R} = \mathbf{Gb} \cdot \Delta I / \sigma_I^2$$

$$\Delta I = 1.591$$

$$\sigma_I^2 = 1.409$$

$$R's = \text{SCN}^- - 5.73 \text{ mg/100g DM, } \text{SMCO} - 0.040 \text{ g/100g DM} \\ \text{and } \text{DOMDY} 0.34 \text{ t/ha.}$$

## Long-Term Selection in Maize

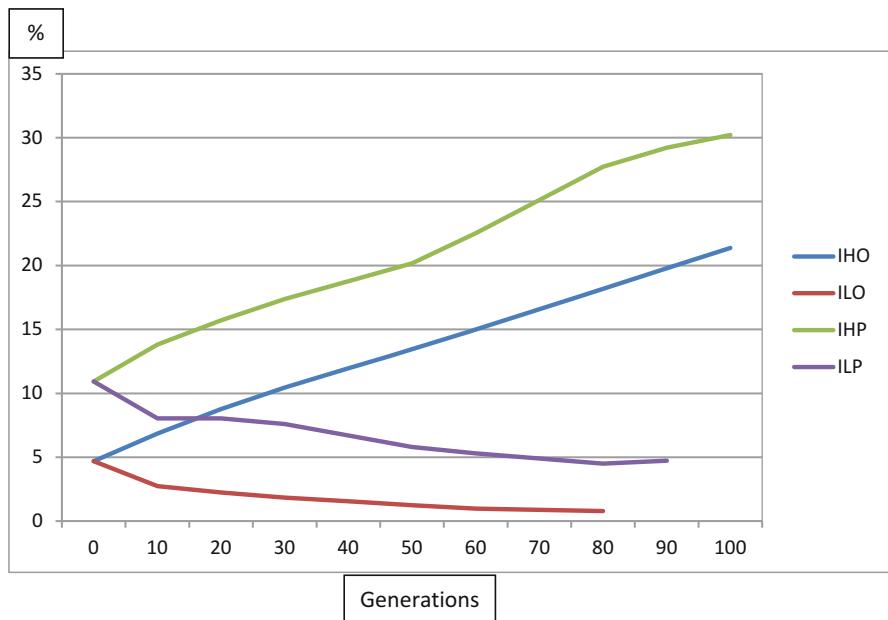
### *Illinois Long Term Selection Experiment*

It is now time to consider the outcome of many cycles of crossing and selection. The best-known long-term selection experiment in plants is for high and low oil and high and low protein content in maize. It was started by Cyril G Hopkins in 1896 in Illinois and has continued for over 100 generations. It is often referred to as the Illinois Long Term Selection Experiment (ILTSE). Volume 24 Part 1 of the Plant Breeding Reviews (Janick 2004) was devoted to reviews of the results of the first 100 generations of selection. The chemical composition of the dry-matter of a typical maize kernel was 4.4 % oil, 11.4 % protein, 82.8 % carbohydrate and 1.4 % ash. The oil and protein contents were considered low for an animal feed and there was increasing demand for corn-oil. The aim of the experiment was to discover the extent to which these contents could be modified, and hence there was scientific interest in selecting for low as well as high levels. The starting material was 163 ears of the open-pollinated corn cultivar 'Burr's White' with a fairly typical average chemical composition. The 24 ears highest in oil, the 12 ears lowest in oil, the 24 ears highest in protein and the 12 ears lowest in protein were selected to initiate the four primary selection lines, and these were grown in separate fields for isolation from each other. Ears from separate plants were analysed for chemical composition and those selected planted ear-to-row in what was in effect a half-sib progeny test. However, there were some differences in selection procedures over the generations, so Dudley and Lambert (2004) summarized the results into four or five segments, as appropriate, where the year to year fluctuations were smoothed by

regression analysis. They also provided the full set of results for each year. The summary shown in Table 10.1 and Fig. 10.2 is derived from their regression analyses so there is smoothing of year to year fluctuations. Over the first ten generations there was a symmetric and reasonably steady response for high and low oil content, with realized heritabilities of 0.50 in both lines (i.e. half of the cumulative selection differentials applied). Likewise there was a symmetric and

**Table 10.1** Selection for high and low oil (IHO and ILO) and protein (IHP and ILP) content in maize (Dudley and Lambert 2004)

Generation	IHO%	ILO%	IHP%	ILP%
0	4.69	4.69	10.93	10.93
10	6.86	2.75	13.82	8.05
20	8.76	2.25	15.72	8.05
30	10.46	1.85	17.37	7.60
40	11.96	1.55	18.77	6.70
50	13.46	1.25	20.17	5.80
60	14.98	0.99	22.53	5.30
70	16.58	0.89	25.13	4.90
80	18.18	0.79	27.73	4.50
90	19.78	—	29.21	4.73
100	21.38	—	30.21	—



**Fig. 10.2** Selection for high and low oil (IHO and ILO) and protein (IHP and ILP) content in maize

reasonably steady response for high and low protein content, but with lower realized heritabilities of 0.17 and 0.24, respectively. Thereafter there was a steady increase in oil and protein content in the high lines with no sign of a selection limit. In contrast in the low lines the decrease slowed and between generations 80 and 90 had ceased. In the low oil line it was not possible to accurately measure lower oil levels, whereas in the low protein line an apparent physiological limit had been reached. As already mentioned, the Plant Breeding Review provides extensive analysis for the interested reader, so here I am just going to highlight a few key conclusions.

### ***Conclusions from Selection Experiment***

At the start of the experiment, the estimates of the average frequencies of alleles ( $q$ ) for high oil and high protein content were 0.20 and 0.24, respectively; the estimates of the number of genes segregating (strictly speaking, effective factors) were 54 and 122, respectively; and the estimates of additive genetic variance ( $\sigma_A^2$ ) were 0.6612 and 0.5032, respectively. Hence the total gains (difference between the mean at 100 generations and at generation 0) in the high and low strains for oil content were  $21.1\sigma_A$  and  $5.2\sigma_A$ , respectively. Likewise the total gains in the high and low strains for protein content were  $27.6\sigma_A$  and  $9.0\sigma_A$ , respectively. Therefore the total gains took the strain means way beyond the extremes seen in the original population of ears. It is possible to explain all of this progress by segregation of the relatively large number of genes, each at a relatively low frequency, in the original population; but mutation cannot be eliminated as a possible source of some of the variation upon which selection continues to operate (Dudley and Lambert 2004). In fact Walsh (2004) suggested that the gain based on mutation variance was expected to exceed that from gain based on residual segregation from the original population after about 46 generations for oil content and 33 for protein content (Dudley and Lambert 2004). Walsh (2004) also concluded that the initial variation for oil content was effectively used by selection so that favourable alleles had a high probability of fixation, whereas some favourable protein alleles were likely to have been lost by chance (i.e. genetic drift). Dudley (2007) summarized the results of QTL analysis as detecting 54–70 QTLs for oil content and 41–50 for protein content. Hence QTL analysis confirmed the large number of genes segregating, but the estimate for protein content was about half that from earlier analysis by the methods of quantitative genetics. Dudley (2007, 2008) also concluded that additive  $\times$  additive epistasis could have contributed to the responses as a result of the effects of QTL alleles changing as their genetic background changed with selection. He also thought that the addition of nitrogen fertilizer from generation 53 could have allowed the expression of alleles for higher protein content.

## Theory of Selection Limit

In the last section we examined the results and analysis of a long term selection experiment, but can we make any useful predictions at the start of a programme about its eventual outcome? In other words, can we predict the selection limit and how quickly it can be achieved?

The total response to selection ( $R_T$ ) that can eventually be achieved relative to the initial additive genetic variation ( $R_T/V_A^{1/2} = R_T/\sigma_A$ ) depends primarily on the number of loci contributing to the variation: the larger the number, the greater the response. This can be demonstrated for a simple additive model (i.e. no dominance or epistasis) with two alleles at each locus, if we assume all loci ( $n$ ) are of equal effect ( $a$ ) and all increasing alleles have the same starting frequency ( $p$ ).

For each locus:

Genotype	$AA$	$Aa$	$aa$
Genotypic value	$a$	0	$-a$
Frequency	$p^2$	$2pq$	$q^2$
Population mean	$M = a(p - q) = a(2p - 1)$		
Population variance	$\sigma_A^2 = 2pqa^2 = 2p(1-p)a^2$		

$R_H$  is total response from the population mean (M) in the high line with all alleles of increasing effect fixed ( $AABBCC\dots NN$ ),  $R_L$  is the total response in the low line with all alleles of decreasing effect fixed ( $aabbcc\dots nn$ ) and  $R_T$  is the difference ( $R_H - R_L$ ).

$$R_H = \Sigma[a - a(2p - 1)] = \Sigma[2a(1 - p)] = 2na(1 - p)$$

$$R_L = \Sigma[-a - a(2p - 1)] = \Sigma[-2ap] = -2nap$$

$$R_T = R_H - R_L = 2na$$

$$R_T^2 = 4n^2a^2$$

$$\sigma_A^2 = 2na^2 p(1 - p)$$

$$R_H^2/\sigma_A^2 = 2n(1 - p)/p$$

$$R_L^2/\sigma_A^2 = 2np/(1 - p)$$

$$R_T^2/\sigma_A^2 = 2n/[p(1 - p)]$$

When  $p = 0.5$ ,  $R_T^2/\sigma_A^2 = 8n$  and  $R_T = (8n)^{1/2}\sigma_A$

In practice, when all of the segregating loci have alleles of different effects and different initial frequencies, there is no simple relationship between the initial additive variance and total response to selection. Nevertheless, our simple and unrealistic model does demonstrate the following points. When  $n$  is small, say 2 giving a  $R_T/\sigma_A$  ratio of 4, a few genes of large effect are quickly fixed within range seen in base population. However, when  $n$  is large, say 32 giving a ratio of 16, or

128 giving 32, many genes of small effect are fixed slowly and outside range of even a large base population. Furthermore, it can be seen that the high and low responses will only be symmetrical when the initial high and low allele frequencies are equal (i.e. 0.5).

Once a long term selection experiment has been completed, the initial average allele frequency can be estimated from the ratio  $R_H/R_L$ .

$$R_H/R_L = (p - 1)/p = 1 - 1/p$$

$p = 1/(1 - R_H/R_L)$ , remembering that  $R_L$  has a negative sign, so that if the absolute values of  $R_H$  and  $R_L$  are used, one has:

$$p = 1/(1 + R_H/|R_L|).$$

Hence when  $R_H = R_L$ ,  $p = 0.5$ , as already mentioned.

The number of loci  $n$  (strictly speaking the number of effective factors  $k$ ) can be estimated from  $R_H$ ,  $R_L$ ,  $p$  and  $\sigma_A^2$ :

$$\begin{aligned} [\frac{1}{2}(R_H - R_L)]^2 &= n^2 a^2 \\ \sigma_A^2 &= 2na^2 p(1 - p) \end{aligned}$$

Hence  $[\frac{1}{2}(R_H - R_L)]^2 / \sigma_A^2 = n/[2p(1 - p)]$  and  $n = [2p(1 - p)][\frac{1}{2}(R_H - R_L)]^2 / \sigma_A^2$

However, to re-iterate, at the start of a breeding programme one cannot predict long-term response simply from knowledge of the base-population variance components. Nevertheless, a combination of some basic theory and additional observations can provide some insight into what can be expected. Thus estimates from QTL analysis of the number of loci segregating, their chromosomal locations and the magnitudes of allelic effects will indicate if a long term response is likely, and if correlated responses (desirable and undesirable) in other traits are expected from linkage and pleiotropy.

## Loss of Desirable Alleles by Genetic Drift

### Theory

We now need to look at another issue that arises from the above considerations and was mentioned in the context of the Illinois Long Term Selection Experiment. If there are a large number of loci segregating, can all of the desirable alleles be fixed by selection, or are some lost by chance? In other words, what is the selection limit (population mean) when at each locus all individuals in a population are homozygous for the same allele? The theory is difficult so a few key results will be

presented, but not their derivation. A detailed account and references to the original literature can be found in the review by Walsh (2004) which was mentioned earlier. If the effective population size ( $N_e$  the effective number of breeding individuals or parents) in a selection programme is modest (less than 100), and a large number of loci are segregating, allele effects are small and allele frequency change is largely governed by drift rather than selection (i.e. weak selection on individual loci). Under these circumstances the infinitesimal model of population genetics provides useful results. This model assumes that there are a very large number of unlinked loci, each of very small additive effect. Alleles of increasing and decreasing effect at segregating loci will slowly be fixed by chance over generations so that eventually no segregating loci are left. The probability of fixation depends on the initial frequency of the allele, but averaged over all loci, the fixation of alleles of increasing and decreasing effect by drift is not expected to result in much change in the population mean. In contrast, the genetic variance decreases over generations. Where (weak) selection is effective in achieving a small increase in the frequency of alleles of increasing effect (upward selection) in the next generation, the population mean is expected to show a small increase in response, but with virtually no change in the population variance as a result of the selection. However, in the next generation the response will be slightly less because there is not quite so much genetic variation available as some has been lost by drift. Starting with the expected decline in additive genetic variance over generations from drift, Robertson (1960) showed that the upper limit for the total response to selection ( $R_{(\max)}$ , the selection limit) is  $2N_e$  times the initial response ( $ih^2\sigma_p = ih\sigma_A$ , where  $h$  and  $\sigma_A$  are the square roots of the heritability and the additive genetic variance of the initial population).

$$R_{(\max)} = 2N_e ih^2 \sigma_p = 2N_e ih\sigma_A.$$

The maximum response to divergent selection is obtained simply by putting  $i$  as the sum of the intensities of selection in the two directions. Robertson (1960) also showed that half of the total response to selection (its half-life) was achieved after  $1.4N_e$ . However, if selection dominates over drift, the initial response to selection is much larger and changes in allele frequencies are much faster, so that  $R_{(\max)}$  and half-life are both overestimated. This is likely to occur with fewer genes of larger effects. Nevertheless, all of the components in the formula for  $R_{(\max)}$  are measurable in the foundation population and hence the formula is a prediction of what might be expected if a very large number of loci were segregating.

A key observation from the above equation is that there is a potential tradeoff between the short-term and long-term response. If  $N$  individuals are selected out of  $M$  assessed, the intensity of selection ( $i$ ) can be increased by reducing  $N$ , and this increases the immediate response to selection ( $ih\sigma_A$ ). However, reducing  $N$  decreases the total response to selection  $R_{(\max)}$ . Robertson (1960) found that the largest total response occurred when half the population is saved, but as  $M$  increases fairly large deviations of  $N/M$  from a half give essentially the same limit. Refinements (i.e. fewer assumptions) of Robertson's model can be found in

the review by Walsh (2004) along with complementary single-locus models in finite populations.

Keightley (2004) has reviewed the role of new mutations in sustaining long-term selection responses. He points out that in many ways the properties of the selection response arising from new mutations resembles that of the long-term response from pre-existing variation, but departures from simple expectations based in the infinitesimal model are exaggerated. Falconer and Mackay (1996) summarize what ought to happen under long-term selection when mutation is taken into account as follows. At first the response comes from the existing additive genetic variance in the base population. The rate of response should diminish gradually as the additive variance is depleted. Then after about 20 generations, new variance from mutated genes begins to contribute to the response, which should diminish more slowly. Eventually all of the genes segregating in the base population should have been brought to fixation by selection or drift and the continuing response depends entirely on the mutations that have accumulated during selection. Although the response could theoretically continue indefinitely at a constant rate, in practice a limit is usually encountered. Possible explanations are discussed by Falconer and Mackay (1996). It is reasonable to question whether these considerations are relevant to plant breeding programmes. Plant breeders probably do not think in terms of more than 20 generations when designing their breeding programmes. They are concerned with the short-term response over say the first ten generations. However, all plant breeding comprises cycles of crossing and selection, usually among the breeders' own advanced germplasm and between this germplasm and new cultivars from other breeding programmes, on the assumption that different desirable genes have been fixed in different potential parents. Breeders are concerned about running out of genetic variation, particularly for yield (yield plateaus), and hence broadening the genetic base of their programmes for continued progress is an issue, as we shall see in Part IV.

### ***Computer Simulation***

As predicting the response to medium-term term selection from theory is difficult, I (Bradshaw 1984) used computer simulations to help me design a kale breeding programme back in the early 1980s. I compared three recurrent selection schemes suitable for kale involving half-sib (HS), full-sib (FS) and selfed (S) families. All combinations of 6, 12 and 24 families selected, out of 120 and 240 families assessed, were investigated for a range of genetic models, but all with 20 unlinked loci of equal effect and with two alleles. Selection was simulated for 20 generations from an initial allele frequency of 0.05 and for 16 generations from an initial frequency of 0.20. The initial allele frequencies were chosen based on 20 or 5 cultivars being included in the foundation population and each contributing unique alleles. With an initial frequency of 0.05 there was a serious loss of desired alleles ranging from 0.31 out of 20 for the HS scheme with 24 out of 240 families

selected to 9.19 for the S scheme with 6 out of 120 families selected. It was concluded that the selection scheme should be chosen to minimize the loss, i.e. 24 out of 240 HS families to be selected. With an initial frequency of 0.20 there were no losses with 12 and 24 families selected in the HS and FS schemes respectively, and the highest loss was 2.88 for the S scheme with 6 out of 120 families selected. It was concluded that a compromise should be sought between the initial response to selection over the first five generations and the loss of desired alleles, e.g. selecting 12 out of 240 FS families. The effective population size of such a population is determined by the relatedness of the 24 parents of the 12 FS families. The simulations guided the design of the actual breeding programme which has now been successfully completed (Bradshaw and Wilson 2012). It should be pointed out that some authors argue that comparisons of schemes for long-term selection should be done under the restriction of equal effective population size (Posselt 2010).

## Cultivar Production: Open-Pollinated Cultivars

At any stage in a population improvement programme the breeder can produce and assess potential new cultivars. In the kale programme mentioned above this was done after five generations by selecting and separately multiplying six half-sib families (Bradshaw and Wilson 2012). Although this can result in a small overall reduction in yield due to inbreeding depression (equal to one quarter of that expected from one generation of self-pollination), it also affords the opportunity to identify superior half-sib families. Thus cultivar Grampian was the one with the highest DOMD content. Had the whole population been considered sufficiently uniform, it could have been multiplied as an open-pollinated cultivar, as was successfully done in producing kale cultivar Caledonian from a clubroot resistant subpopulation. As open-pollinated cultivars are genetically variable, maintaining them true to type is very important. This involves both rousing out any off-types and positive selection for the cultivar description, which in the European Union is based on the definitive stock submitted for DUS (Distinctness, Uniformity and Stability) assessment, as part of National Listing, and for Plant Breeders' Rights. Furthermore, the population size of the first stage of multiplication must be sufficiently large to avoid inbreeding depression. In the kale programme, pre-basic seed was produced in 12 m long polythene tunnels from 100 plants selected from a field-grown plot sown with the breeder's stock of seed. The 100 plants were officially inspected at this stage and again at flowering. Pre-basic seed was then multiplied in a suitable isolated field to produce basic seed which in turn was used to produce certified seed for sale, both generations having been subject to official inspections by the certifying authority. Seed marketing regulations, including purity and quality standards, can be found on the website of the certifying authority, as explained in Chap. 2. An open-pollinated cultivar is maintained by periodically repeating the procedure just described, starting with the definitive stock, or material derived from

this stock which has been shown to be true to type. Breeders store seed reserves under appropriate conditions to limit the number of generations (from the definitive stock) in seed multiplication and to minimize the risk of failing certification.

There is no reason why the improved population should not be used as the starting material for an inbreeding and selection programme aimed at hybrid cultivars. In fact the production of hybrid cultivars has been the driving force behind much of the plant breeding done during the twentieth century, and now in the twenty-first century, as we shall see in later chapters.

## Cultivar Production: Synthetic Cultivars

The production of synthetic cultivars from improved populations has proved particularly popular in perennial forage crops and amenity grasses, because their superiority over open-pollinated cultivars makes the additional work involved worthwhile. A typical scheme can be found in the review by Posselt (2010), together with a summary of the theory. More detailed theoretical considerations can be found in the book by Wricke and Weber (1986). Synthetic cultivars are maintained and produced from relatively few component parents. The parents of a synthetic cultivar can be partially or completely inbred lines, noninbred heterozygous plants or clones, families, or whole populations. Here we are going to concentrate on clones. In contrast, open-pollinated cultivars are maintained and produced from the whole improved population.

The first stage of production is the selection of potential parents from the improved population. In perennial forage crops and amenity grasses individual plants can be cloned and evaluated as clonal rows. Commonly at least one third of these are discarded because of their inferior performance. The surviving clones are then testcrossed and their progenies evaluated to determine the general combining abilities of the clones. This means trying to ensure that each clone is pollinated by a random sample of pollen from the other clones. Ideally, each clone should also be self-pollinated and the selfed progenies evaluated; but this may prove difficult because of gametophytic self-incompatibility in most forage crops and amenity grasses. We shall see shortly that reliable estimates of general combining ability, and ideally selfed-progeny performance, are required for predicting the best synthetics to make and assess as potential cultivars. Testing should take place in more than one location for more than 1 year to assess the extent of genotype  $\times$  environment interactions, and with grasses should involve more than one cut. The final choice of parents is based on synthetic prediction, but also taking DUS requirements into account. The parents are the Syn-0 generation. Synthesis is commonly done from the chosen parental clones; although with improved seed storage facilities, remnant seed from the testcrosses is an attractive alternative, despite the loss of (male) parental control. A complete (diallel) set of crosses without selfs among the parents produces the Syn-1 generation. With self-incompatible parents this can be achieved either by open-pollination or controlled

crossing without emasculation. Equal quantities of seed of each cross are mixed and then multiplied by open-pollination to produce the Syn-2 generation. This is the earliest generation for entry into National Listing Trials (DUS and VCU). Further seed multiplication is undertaken for commercialization, certified seed commonly being Syn-4. The parental clones need to be maintained (*in situ* or *in vitro*) for reconstitution of the synthetic as needed, but reserve Syn-1 and Syn-2 seed is likely to be held in cold storage.

### ***Prediction of Performance of Synthetic Cultivars: A Simple Diploid Example***

A simple example will provide the reader with insight into the theory which can be found in the literature and which is summarized by Posselt (2010). It is one of the more difficult areas of theoretical plant breeding.

Let us start by considering a synthetic cultivar produced from just four parents, and let us consider a single locus (*A*) with two alleles (*A* and *a*). Let us suppose that the four parents (Syn-0) have the following genotypes and genetic values.

P <sub>1</sub>	P <sub>2</sub>	P <sub>3</sub>	P <sub>4</sub>
AA	AA	<i>Aa</i>	<i>aa</i>
+ <i>a</i>	+ <i>a</i>	<i>d</i>	- <i>a</i>

The allele frequencies are  $p = 5/8$  and  $q = 3/8$  for '*A*' and '*a*', and the mean of the four parents is  $\frac{1}{4}a + \frac{1}{4}d$ .

The six possible crosses (Syn-1) between the four parents are as follows.

C <sub>12</sub>	C <sub>13</sub>	C <sub>14</sub>	C <sub>23</sub>	C <sub>24</sub>	C <sub>34</sub>
AA	( $\frac{1}{2}AA + \frac{1}{2}Aa$ )	<i>Aa</i>	( $\frac{1}{2}AA + \frac{1}{2}Aa$ )	<i>Aa</i>	( $\frac{1}{2}Aa + \frac{1}{2}aa$ )
+ <i>a</i>	$\frac{1}{2}(a+d)$	<i>d</i>	$\frac{1}{2}(a+d)$	<i>d</i>	$\frac{1}{2}(-a+d)$

The allele frequencies are  $p = 5/8$  and  $q = 3/8$ , the genotype frequencies are 4/12 (*AA*), 7/12 (*Aa*), and 1/12 (*aa*), and the mean of the six crosses is  $\frac{1}{4}a + (7/12)d$ . Hence the allele frequencies are the same, the genotype frequencies have changed, and the mean has increased by  $(1/3)d$ . The genotype frequencies are not in Hardy-Weinberg equilibrium.

The six crosses now make equal contributions to the next generation (Syn-2) through random mating. In other words, the male and female gametes each have allele frequencies of  $p = 5/8$  and  $q = 3/8$ , so the result is a population in Hardy-Weinberg equilibrium with genotype frequencies 25/64 (*AA*), 30/64 (*Aa*), and 9/64 (*aa*), and the population mean is  $\frac{1}{4}a + (15/32)d$ . Hence compared with Syn-1, the population mean has decreased by  $(11/96)d$ . As the population is now in Hardy-Weinberg equilibrium, there are no changes in subsequent generations and

Syn-2 = Syn- $\infty$  (infinity or equilibrium so Syn-e also used). In most forage crops a slight yield decrease is observed from Syn-1 to Syn-2 (Posselt 2010).

The Syn- $\infty$  population can be produced in another way, namely making all possible crosses (C) between the four parents (P), including reciprocal crosses, and all possible parent by parent crosses which in our example are the four selfs (S). Thus in an outbreeding crop the selfs are the S<sub>1</sub> generation.

	P <sub>1</sub>	P <sub>2</sub>	P <sub>3</sub>	P <sub>4</sub>
P <sub>1</sub>	S <sub>11</sub>	C <sub>12</sub>	C <sub>13</sub>	C <sub>14</sub>
P <sub>2</sub>	C <sub>21</sub>	S <sub>22</sub>	C <sub>23</sub>	C <sub>24</sub>
P <sub>3</sub>	C <sub>31</sub>	C <sub>32</sub>	S <sub>33</sub>	C <sub>34</sub>
P <sub>4</sub>	C <sub>41</sub>	C <sub>42</sub>	C <sub>43</sub>	S <sub>44</sub>

In our example the 16 progenies are as follows, from which the genotype frequencies can be confirmed as the Hardy-Weinberg ones given above. (There are no reciprocal differences: C<sub>14</sub> = C<sub>41</sub> and C<sub>14</sub> + C<sub>41</sub> = 2C<sub>14</sub>.)

Parents	AA	AA	Aa	aa
AA	AA	AA	½AA + ½Aa	Aa
AA	AA	AA	½AA + ½Aa	Aa
Aa	½AA + ½Aa	½AA + ½Aa	¼AA + ½Aa + ¼aa	½Aa + ½aa
aa	Aa	Aa	½Aa + ½aa	aa

With more than one locus, the Syn-1 generation will not be in gametic phase equilibrium, and this equilibrium will only be reached gradually over subsequent generations. Hence in the presence of epistasis, the mean of the synthetic will continue to change after Syn-2, whereas in the absence of epistasis, the Syn-2 mean is the equilibrium mean and our simple example has general applicability. The reader is referred to Wricke and Weber (1986) for more advanced theory.

### ***Prediction of Performance of Synthetic Cultivars: Some More General Diploid Results***

In the above example we considered four parents. If there are  $n$  parents, then crossing in all combinations and making all selfs results in  $n^2$  progenies of which  $n$  are selfs and  $n(n - 1)$  are crosses. If the mean (e.g. yield) of the crosses is  $C$  and the mean of the selfs is  $S$ , then the equilibrium yield of the synthetic is:

$$Y = [n(n - 1)C + nS]/n^2 = C - (C - S)/n$$

Hence with a 4-parent synthetic,  $Y = C - (C - S)/4$ , where  $C$  is the mean of the six different crosses (assuming reciprocals have same yield) and  $S$  is the mean of the four selfs. In our example above,  $C = \frac{1}{4}a + (7/12)d$ ,  $S = \frac{1}{4}a + (1/8)d$  and  $Y = \frac{1}{4}a + (15/32)d$ , which is Syn-∞. In addition to this 4-parent synthetic, the four parents could be used to produce four 3-parent synthetics and six 2-parent synthetics. Which is the best synthetic? With just four parents we could produce all 11 synthetics and find out by trialling them. However, the number of possible synthetics is  $2^n - n - 1$  (Wricke and Weber 1986). So the number increases from 11 with four parents, to 1013 with 10 parents to 1,048,555 with 20 parents, and so on. We need a method to predict the performance of synthetics without making them all! With 20 parents, we could make all 190 crosses and 20 selfs and evaluate them, but if there were 100 potential parents these numbers would rise to 4950 crosses and 100 selfs. So initially, let's return to our four parents and see what can be done.

We saw above that the yield ( $Y$ ) of a 4-parent synthetic is:

$$Y_{(4-\text{syn})} = C - (C - S)/4 [= C_{1234} - (C_{1234} - S_{1234})/4]$$

Now let us consider the four possible 3-parent synthetics and use subscripts to denote their parents. Their yields are:

$$\begin{aligned} Y_{123} &= C_{123} - (C_{123} - S_{123})/3 \\ Y_{124} &= C_{124} - (C_{124} - S_{124})/3 \\ Y_{134} &= C_{134} - (C_{134} - S_{134})/3 \\ Y_{234} &= C_{234} - (C_{234} - S_{234})/3 \end{aligned}$$

where  $Y_{123}$  is the equilibrium yield of the 3-parent synthetic using parents  $P_1$ ,  $P_2$  and  $P_3$ , and  $C_{123} = (C_{12} + C_{13} + C_{23})/3$ , and  $S_{123} = (S_{11} + S_{22} + S_{33})/3$ , etc.

Now if we work out the mean of these four 3-parent synthetics we find that:

$$\bar{Y}_{(3-\text{syn})} = C_{1234} - (C_{1234} - S_{1234})/3$$

Now let us consider the six possible 2-parent synthetics. Their yields are:

$$\begin{aligned} Y_{12} &= C_{12} - (C_{12} - S_{12})/2 \\ Y_{13} &= C_{13} - (C_{13} - S_{13})/2 \\ Y_{14} &= C_{14} - (C_{14} - S_{14})/2 \\ Y_{23} &= C_{23} - (C_{23} - S_{23})/2 \\ Y_{24} &= C_{24} - (C_{24} - S_{24})/2 \\ Y_{34} &= C_{34} - (C_{34} - S_{34})/2 \end{aligned}$$

Now if we work out the mean of these six 2-parent synthetics we find that:

$$\bar{Y}_{(2-\text{syn})} = C_{1234} - (C_{1234} - S_{1234})/2$$

More generally, if we start with an  $n$ -parent synthetic and consider all of the  $k$ -parent synthetics that can be produced, their mean will be:

$$\bar{Y}_{(k-syn)} = \mathbf{C} - (\mathbf{C} - \mathbf{S})/\mathbf{k}$$

where  $C$  is the mean of the  $n(n - 1)$  crosses and  $S$  is the mean of the  $n$  selfs.

As outbreeding crops display inbreeding depression,  $C$  will be greater than  $S$  and  $\bar{Y}_{(k-syn)}$  will increase as  $k$  increases. But this doesn't mean that the  $n$ -parent synthetic is the best one because there is only one  $n$ -parent synthetic but  $\frac{1}{2}n(n - 1)$  2-parent synthetics, for example.

We now need to consider how to predict the best synthetic of size  $k$ .

There is another way of looking at the yields of the crosses. Let us go back to our 4-parent example. The yields of the six crosses can be modelled as follows:

$$\begin{aligned} C_{12} &= C + gca_1 + gca_2 \\ C_{13} &= C + gca_1 + gca_3 \\ C_{14} &= C + gca_1 + gca_4 \\ C_{23} &= C + gca_2 + gca_3 \\ C_{24} &= C + gca_2 + gca_4 \\ C_{34} &= C + gca_3 + gca_4 \end{aligned}$$

where  $C$  is the mean of the six crosses and the  $gca$ 's (general combining abilities) are departures from this mean that therefore sum to zero ( $gca_1 + gca_2 + gca_3 + gca_4 = 0$ ). If we sum all of the crosses with a common parent, say parent 1, we find that we have an estimate of  $gca_1$ :

$$\begin{aligned} C_{12} + C_{13} + C_{14} &= 3C + 3gca_1 + gca_2 + gca_3 + gca_4 = 3C + 2gca_1 \\ (C_{12} + C_{13} + C_{14})/3 - C &= \frac{2}{3}gca_1 \end{aligned}$$

In practice, if we pollinate parent 1 with a mixture of pollen from all of the other parents and assess the resulting half-sib family along with all of the other possible half-sib families, we can estimate the general combining abilities as departures from the mean  $C$ , which is the mean of all of the half-sib families. If the parents are all self-incompatible, natural pollination can be used to produce the half-sib families. Differences can occur between the actual yields of the six crosses and those predicted from the general combining abilities as a result of specific combining abilities unique to each cross. Estimates of specific combining abilities are difficult to obtain when the number of potential parents is large (one cannot make a complete diallel set of crosses), but fortunately using only general combining abilities has proved adequate for practical purposes.

Let us look at the yield ( $Y_{12}$ ) of a 2-parent synthetic:

$$Y_{12} = C_{12} - (C_{12} - S_{12})/2 = \frac{1}{2}C_{12} + \frac{1}{2}S_{12} = \frac{1}{2}(C + gca_1 + gca_2) + \frac{1}{2}S_{12}$$

The mean of all six 2-parent synthetics is:

$$\bar{Y}_{(2-\text{syn})} = C - (C - S)/2 = \frac{1}{2}C + \frac{1}{2}S, \text{ where } C = C_{1234} \text{ and } S = S_{1234}$$

$$\text{Hence } Y_{12} - \bar{Y}_{(2-\text{syn})} = \frac{1}{2}(C + gca_1 + gca_2) + \frac{1}{2}S_{12} - \frac{1}{2}C - \frac{1}{2}S = \frac{1}{2}(gca_1 + gca_2) + \frac{1}{2}(S_{12} - S)$$

and therefore

$$Y_{12} = \bar{Y}_{(2-\text{syn})} + \frac{1}{2}(gca_1 + gca_2) + \frac{1}{2}(S_{12} - S)$$

Now let us look at the yield ( $Y_{123}$ ) of a 3-parent synthetic:

$$Y_{123} = C_{123} - (C_{123} - S_{123})/3 = \frac{2}{3}C_{123} + \frac{1}{3}S_{123}$$

$$\begin{aligned} C_{123} &= (C_{12} + C_{13} + C_{23})/3 = (3C + 2gca_1 + 2gca_2 + 2gca_3)/3 \\ &= C + \frac{2}{3}(gca_1 + gca_2 + gca_3) \end{aligned}$$

Therefore

$$Y_{123} = \frac{2}{3}C_{123} + \frac{1}{3}S_{123} = \frac{2}{3}C + \left(\frac{2}{3}\right)\left(\frac{2}{3}\right)(gca_1 + gca_2 + gca_3) + \frac{1}{3}S_{123}$$

The mean of all four 3-parent synthetics is:

$$\bar{Y}_{(3-\text{syn})} = C - (C - S)/3 = \frac{2}{3}C + \frac{1}{3}S$$

Hence

$$\begin{aligned} Y_{123} - \bar{Y}_{(3-\text{syn})} &= \frac{2}{3}C + \left(\frac{2}{3}\right)\left(\frac{2}{3}\right)(gca_1 + gca_2 + gca_3) + \frac{1}{3}S_{123} - \frac{2}{3}C \\ &\quad - \frac{1}{3}S \\ &= \left(\frac{2}{3}\right)\left(\frac{2}{3}\right)(gca_1 + gca_2 + gca_3) + \frac{1}{3}(S_{123} - S) \end{aligned}$$

Therefore

$$Y_{123} = Y_{(3-\text{syn})} + \binom{2}{3} \binom{2}{3} (gca_1 + gca_2 + gca_3) + \frac{1}{3} (S_{123} - S)$$

There is just one 4-parent synthetic, but it is worth doing the same calculation.

$$\begin{aligned} Y_{1234} &= \frac{3}{4}C_{1234} + \frac{1}{4}S_{1234} \\ C_{1234} &= (C_{12} + C_{13} + C_{14} + C_{23} + C_{24} + C_{34})/6 \\ &= (6C + 3gca_1 + 3gca_2 + 3gca_3 + 3gca_4)/6 \\ &= C + \frac{1}{2}(gca_1 + gca_2 + gca_3 + gca_4) \end{aligned}$$

Therefore

$$Y_{1234} = \frac{3}{4}C_{1234} + \frac{1}{4}S_{1234} = \frac{3}{4}C + \frac{3}{4}\frac{1}{2}(gca_1 + gca_2 + gca_3 + gca_4) + \frac{1}{4}S_{1234}$$

The mean of the one 4-parent synthetic is:

$$Y_{(4-\text{syn})} = \frac{3}{4}C + \frac{1}{4}S$$

Hence

$$\begin{aligned} Y_{1234} - \bar{Y}_{(4-\text{syn})} &= \frac{3}{4}C + \frac{3}{4}\frac{1}{2}(gca_1 + gca_2 + gca_3 + gca_4) \\ &\quad + \frac{1}{4}S_{1234} - \frac{3}{4}C - \frac{1}{4}S \\ &= \frac{3}{4}\frac{1}{2}(gca_1 + gca_2 + gca_3 + gca_4) + \frac{1}{4}(S_{1234} - S) \end{aligned}$$

Therefore

$$Y_{1234} = \bar{Y}_{(4-\text{syn})} + \frac{3}{4}\frac{1}{2}(gca_1 + gca_2 + gca_3 + gca_4) + \frac{1}{4}(S_{1234} - S)$$

However, as there is only one 4-parent synthetic, the *gca*'s sum to zero and  $S_{1234} = S$ .

Hence  $Y_{1234} = \bar{Y}_{(4-\text{syn})}$ , and all is well!

If the numbering of the parents is such that  $gca_1 > gca_2 > gca_3 > gca_4$  etc., and if there is little variation over parents in the effect of selfing (i.e. departures from  $S$  can be ignored), then the best synthetic of size  $k$  will be the one with parents 1 to  $k$  out of the  $n$  available.

$$Y_{1\dots k} = \bar{Y}_{(k-\text{syn})} + [2(k-1)/k^2](gca_1 + gca_2 + \dots + gca_k) + (1/k)(S_{1\dots k} - S)$$

Where  $\bar{Y}_{(k-\text{syn})} = C - (C - S)/k$

The best synthetic can therefore be predicted from the general combining abilities and  $S_1$  performance of the potential parents. These are sometimes combined to estimate what is either called the general varietal ability (GVA) or the general synthesizing ability (GSA) of a parent (Wricke and Weber 1986), but they need not concern us further in this book. If the  $S_1$  generation is not available, an estimated value of  $S$  ( $S = x\%$  of  $C$ ) can be used from other work and all  $(S_{1\dots k} - S)$  set equal to zero. This can prove adequate for practical purposes when  $k$  is larger than about 5, but actual  $S_1$  values are best.

As already mentioned, as  $k$  increases, so does  $\bar{Y}_{(k-\text{syn})}$ , and its maximum value is when  $k = n$ . In contrast, the sum of the  $gca$ 's will then be zero so that their sum has its minimum value. When  $k = 2$ ,  $\bar{Y}_{(k-\text{syn})}$  has its minimum value, but the  $gca$ 's will have their maximum value of  $gca_1 + gca_2$ . Hence we can anticipate an optimum intermediate value of  $k$ . A large number of experimental investigations on the optimum number of parents can be found in the literature and have been summarized by Becker (1988). In most experiments the optimum number of components was 5, and the use of more than 10 clones was never recommended. Posselt (2010) concluded that with greater inbreeding depression, higher numbers of synthetic parents are required, but that the optimum was still likely to be in the range 7–11. However, DUS restrictions may reduce the number of parents to between 4 and 10 (Humphreys et al. 2010).

### ***Complications in Autotetraploids***

As some forage crops are autotetraploids (alfalfa, red clover and a number of grasses), we shall briefly consider the performance of synthetic cultivars under tetrasomic inheritance. Once again the breeder needs to seek the best compromise between few parents to maximize their general combining abilities and many parents to minimize the effects of inbreeding. The general combining abilities of all potential parents and the effects of selfing them provides the relevant information, but the prediction equations are more complicated and assumptions have to be made about the effects of inbreeding. Below it will be seen that inbreeding depends on  $h$ ,  $v$  and  $w$ , not simply  $h$ . Complexities arise because diploid gametes can transmit diallelic interactions (dominance effects) and inbreeding effects (two alleles identical by descent) to the next generation. One generation of random mating does not achieve the single locus equilibrium of genotype frequencies, nor does it remove the effects of inbreeding. The complications that arise can be seen in the following simple example.

Let us produce a synthetic cultivar from four parents, and let us consider a single locus ( $A$ ) with two alleles ( $A$  and  $a$ ). Let us suppose that the four parents (Syn-0) have the following genotypes and genetic values.

P <sub>1</sub>	P <sub>2</sub>	P <sub>3</sub>	P <sub>4</sub>
AAAA	AAAa	AAaa	Aaaa
+2a	+a + 3h + v + w	+4h	-a + 3h - v + w

The allele frequencies are  $p = 5/8$  and  $q = 3/8$  for ' $A$ ' and ' $a$ ', and the mean of the four parents is  $\frac{1}{2}(a + 5h + w)$ .

The six possible crosses (Syn-1) between the four parents are as follows.

C <sub>12</sub>	C <sub>13</sub>	C <sub>14</sub>	C <sub>23</sub>	C <sub>24</sub>	C <sub>34</sub>
$\frac{1}{2}AAAA$	(1/6)AAAA		(1/12)AAAA		
$\frac{1}{2}AAAa$	(4/6)AAAa	$\frac{1}{2}AAa$	(5/12)AAa	$\frac{1}{4}AAa$	(1/12)AAa
	(1/6)AAaa	$\frac{1}{2}Aaa$	(5/12)AAaa	$\frac{1}{2}AAaa$	(5/12)AAaa
			(1/12)Aaaa	$\frac{1}{4}Aaaa$	(5/12)Aaaa
					(1/12)aaaa

$$C_{12} = +1\frac{1}{2}a + 1\frac{1}{2}h + \frac{1}{2}v + \frac{1}{2}w$$

$$C_{13} = +a + 2\frac{2}{3}h + \frac{2}{3}v + \frac{2}{3}w$$

$$C_{14} = +\frac{1}{2}a + 3\frac{1}{2}h + \frac{1}{2}v + \frac{1}{2}w$$

$$C_{23} = +\frac{1}{2}a + \frac{19}{6}h + \frac{1}{3}v + \frac{1}{2}w$$

$$C_{24} = 3\frac{1}{2}h + \frac{1}{2}w$$

$$C_{34} = -\frac{1}{2}a + \frac{19}{6}h - \frac{1}{3}v + \frac{1}{2}w$$

The allele frequencies are  $p = 5/8$  and  $q = 3/8$ .

The genotype frequencies are 9/72 (AAAA), 29/72 (AAAa), 24/72 (AAaa), 9/72 (Aaaa) and 1/72 (aaaa).

The mean of the six crosses is  $(6/12)a + (35/12)h + (5/18)v + (19/36)w$ .

Hence the allele frequencies are the same, the genotype frequencies have changed, and the mean has increased by  $(5/12)h + (5/18)v + (1/36)w$ . The genotype frequencies are not in single locus equilibrium.

When these six crosses are random mated [allele frequencies AA (55/144), Aa (70/144) and aa (19/144)], their progenies form the next generation (Syn-2) with the following composition: 0.146 (AAAA), 0.371 (AAAa), 0.337 (AAaa), 0.128 (Aaaa) and 0.017 (aaaa).

**Table 10.2** Frequencies of genotypes in four-parent synthetic

	<i>AAAA</i>	<i>AAa</i>	<i>Aaaa</i>	<i>Aaaa</i>	<i>aaaa</i>
Parents	0.250	0.250	0.250	0.250	0.000
Selfs	0.319	0.181	0.250	0.181	0.069
Syn-1	0.125	0.403	0.333	0.125	0.014
Syn-2	0.146	0.371	0.337	0.128	0.017
Syn-2*	0.174	0.347	0.313	0.139	0.028
Syn-∞	0.153	0.366	0.330	0.132	0.020

Unlike in diploids, these genotype frequencies are not in single locus equilibrium (Syn-∞), although they are not that far away, as can be seen in Table 10.2. Furthermore, the population (Syn-2\*) produced by making all possible crosses between the four parents (P), including reciprocal crosses, and all possible parent by parent crosses (the four selfs), differs in composition from Syn-2, but likewise is not in equilibrium. This equilibrium will only be reached gradually over subsequent generations and the mean of the synthetic will continue to change after Syn-2.

The mean of four selfs is	$(6/12)a + (50/24)h + (0/72)v + (26/72)w.$
The mean of six crosses is	$(6/12)a + (70/24)h + (20/72)v + (38/72)w.$
The mean of Syn-2* is	$(6/12)a + (65/24)h + (15/72)v + (35/72)w.$
The mean of Syn-2* is	$(0.5)a + (2.708)h + (0.208)v + (0.486)w.$
The mean of Syn-2 is	$(0.5)a + (2.845)h + (0.243)v + (0.499)w.$
The mean of Syn-∞ is	$(0.5)a + (2.812)h + (0.234)v + (0.498)w.$

In order to predict the mean of Syn-∞ from the means of crosses and parents, or parents' selfed, one needs to know the inbreeding coefficient of the parents or to reasonably assume that it is zero. Details can be found in the books by Wricke and Weber (1986) and Gallais (2003). In practice, a breeder could adopt the simple approach of determining the general combining abilities of potential parents and then making and testing synthetics comprising the 2, 4, 6, 8, 10 and 12 parents with the highest general combining abilities, to see which is best. Finally, it must be remembered that if Syn-2 seed is sold, the mean of Syn-2 measures the value of the synthetic correctly and not the mean at equilibrium.

## Faba Bean Synthetics

The following brief account is taken from the review by Link and Ghaouti (2012). Faba bean (*Vicia faba*) is an important nitrogen-fixing crop with protein-rich seeds (about 30 %). Dry faba bean seeds are mainly used as a component of animal feed (ruminants, monogastrics and birds), but are also a staple food in some countries. The main producers are China, Ethiopia and European countries

such as France and the UK. Faba beans for animal feed are commonly grown as an arable crop and are harvested by combine. In temperate regions, both spring and winter cultivars are grown. Faba bean is a diploid ( $2n=2x=12$ ) species which is pollinated by insects, particularly bees, and involves a mechanical tripping mechanism that stimulates the stigma. The degree of cross-pollination is about 50 % and highly variable, as we learnt in the previous chapter; and this mixed mating system raises issues over the best type of cultivar. Although hybrid cultivars to exploit heterosis have been produced in similar crops such as oilseed rape, this has not proved possible in faba beans and hence synthetic cultivars are an attractive alternative. Inbred lines are produced by manual tripping and self-pollination in isolation cages that prevent access by pollinators. The lines are then tested for their general combining ability in progenies produced through natural pollination in a polycross. In this way the progenies reflect parental line differences in degree of cross-pollination as well as *per se* performance and combining ability. A synthetic cultivar will commonly be initiated from four to eight lines. As long as cross-fertilization is less than 100 %, maximum heterosis is not achieved in Syn-1. With 67 % self-fertilization, it takes until Syn-4 when 50 % of possible heterosis can be realized, whereas with 33 % self-fertilization, it is achieved in Syn-2 and 80 % of possible heterosis is realized.

# **Chapter 11**

## **Clonal Cultivars from Multistage Multitrait Selection**

### **Introduction**

The idea was introduced in the last chapter that all plant breeding comprises recurrent cycles of hybridizations followed by selection. It was mentioned that many of the breeding schemes for vegetatively propagated crops are in fact simple mass selection, where the selection is multitrait and multistage over clonal generations. We are therefore going to use the breeding of vegetatively propagated crops to introduce ideas about multitrait, multistage selection. These crops include root and tuber crops, tropical plantation crops, small fruits and tree fruits and nuts. Later we shall look at some examples of actual breeding programmes. Most programmes have traditionally used artificial hybridization to generate genetic variation in large seedling populations, where each seedling is a potential new cultivar. The selection process then takes place over a number of clonal generations in which the surviving clones are evaluated more extensively as increasing amounts of planting material become available. There is just no way that tens of thousands of clones can be multiplied and assessed in replicated yield trials at many sites for a number of years. We will look briefly at the choice of parents, the number and choice of crosses to be made and the size of progenies, before concentrating on the theory and practice of selection within progenies. A consequence of multistage selection is that the time between cycles of hybridization is increased and as a consequence further progress is delayed. Another consequence is that breeders still like to make crosses each year and to have each clonal generation represented by a set of crosses. This results in overlapping generations. Nevertheless, when one looks at the pedigree of a successful cultivar one sees that it is the product of a particular set of cycles of hybridization and selection. Whilst the differences between the clones produced each cycle are due to additive and non-additive genetic variation, it is the additive variation that is made use of long term over the cycles of hybridization and

selection. We will finish the chapter with a brief look at clonal cultivars produced by apomixis, a form of asexual reproduction that gives the appearance of sexual reproduction: “seed without sex”.

## Hybridization Strategy

The design of a hybridization strategy in a breeding programme is straightforward when one knows the sources and number of alleles that need to be combined in a new cultivar to achieve ones objectives. The selection process is also straightforward if the alleles can be uniquely recognized at the DNA level or through their expression. Currently, however, we still need an overall crossing strategy based on quantitative traits where the number of desired alleles is unknown and individual alleles are not recognized. Here a strategy will be outlined along with the reasoning involved. The key points are as follows, and are based on a paper by Simmonds (1996). (Some vegetatively propagated crops are autopolyploids, but this does not affect the arguments. Potatoes, for example, are autotetraploids.)

### *Choice of Hybridizations*

The initial choice of pair crosses needs to be based on differences in mid-parent (phenotypic) values for the set of desired traits, as this is the best predictor of progeny means. How good a prediction, depends on the narrow-sense heritabilities of the traits. We can get a simple guide by considering pairs of clones taken at random from an equilibrium population and using them in pair crosses to create a new population of offspring progenies. The predicted offspring means are determined from their regression on the mid-parent values, the slope of which is estimated as the covariance of offspring and mid-parent values divided by the variance of mid-parent values. For autotetraploid potato, the covariance is  $(1/2)\sigma_A^2 + (1/6)\sigma_D^2$  and the variance is  $(1/2)\sigma_A^2 + (1/2)\sigma_D^2 + (1/2)\sigma_T^2 + (1/2)\sigma_Q^2 + (1/2)\sigma_e^2/r$ , where  $\sigma_A^2$ ,  $\sigma_D^2$ ,  $\sigma_T^2$  and  $\sigma_Q^2$  are the variances due to the main effects of alleles and the diallelic, triallelic and tetra-allelic interactions between alleles  $A_i$ ,  $A_j$ ,  $A_k$  and  $A_l$  at a locus ‘A’, summed over loci, and  $\sigma_e^2/r$  is the plot to plot environmental variation divided by the number of replicates  $r$  of each clone (whether parent or offspring) (Bradshaw 1994b). Desired traits should be combined into an index of overall merit, but are often considered individually by the breeder. It is clearly important to have as much information as possible about potential parents. Breeders usually find that they have generated substantial genetic variance both between and within progenies (full-sib families) and that the distributions of variation are approximately normal. Simmonds (1996) provided evidence from five data sets (two in potatoes, two in sugarcane, and one in rubber) for the normality of distribution of progeny (family)

means. As one cannot predict within cross variances with any accuracy from the parents, the assumption is made that all crosses have the same within cross variance because there is no correlation between mean and variance.

### ***Discarding Inferior Progenies***

Simmonds (1996) demonstrated statistically that the proportion of segregates within crosses superior to standards (controls) falls away quickly with increasing rank (best to worse) of crosses, where realistic standards (best cultivars currently available) have means close to the means of the best crosses. Hence the best few crosses contribute most of the superior segregates and there is no need to make and evaluate other crosses from parents with inferior mid-parent values, unless these are a poor predictor of offspring means due to non-additive genetic variance. Among those crosses that are made, once inferior ones are recognized, whole progenies can be discarded as unlikely to contain superior segregates. This can be done with reasonable precision by assessing whole progenies (without identifying individual clones) in plots in replicated trials, ideally across sites and seasons, and in specialized disease tests. Then clonal assessment can begin within the surviving progenies. Plant breeders commonly find that large between progeny differences do occur and that outstanding new cultivars nearly always emerge from exceptionally good parental combinations.

### ***Number of Crosses and Progeny Sizes***

Where many crosses are made because of little knowledge of potential parents in order to maintain the momentum of a programme, the number of crosses and progeny sizes should be roughly equal, on the assumption of equal amounts of genetic variation. Again in our simple genetic model for autotetraploid potatoes, the genetic variation between and within full-sib families (progenies) is  $(1/2)\sigma_A^2 + (2/9)\sigma_D^2 + (1/12)\sigma_T^2 + (1/36)\sigma_Q^2$  (between) and  $(1/2)\sigma_A^2 + (7/9)\sigma_D^2 + (11/12)\sigma_T^2 + (35/36)\sigma_Q^2$  (within) (Bradshaw 1994b). In the absence of non-additive genetic variance these are equal, but in its presence, the within is greater than the between.

### ***Intensity of Selection***

Too-many crosses and too-large progenies are wasteful of resources because of the relationship between the intensity of selection ( $i$ ) and population size. From the properties of the normal distribution, the proportions of a population which

are expected to exceed 1.28 ( $i = 1.755$ ), 2.33 (2.665), 3.09 (3.367) and 3.72 (3.960) standard deviations from the mean are 1 in 10, 100, 1000 and 10,000, respectively (Falconer and Mackay 1996). In other words, the intensity of selection is proportional to the logarithm of population size and hence is a declining function of numbers. A sample size of 200 crosses each with progeny size of 200 ( $200 \times 200 = 40,000$  seedlings) is large enough to contain the top five crosses and the top five clones within each with a mean greater than two standard deviations from the sample mean.

### **Theoretical Example**

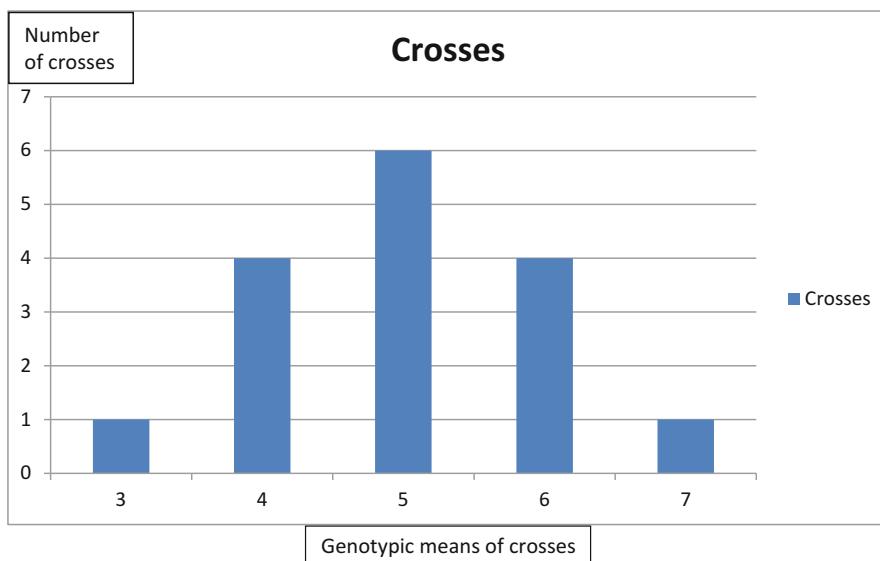
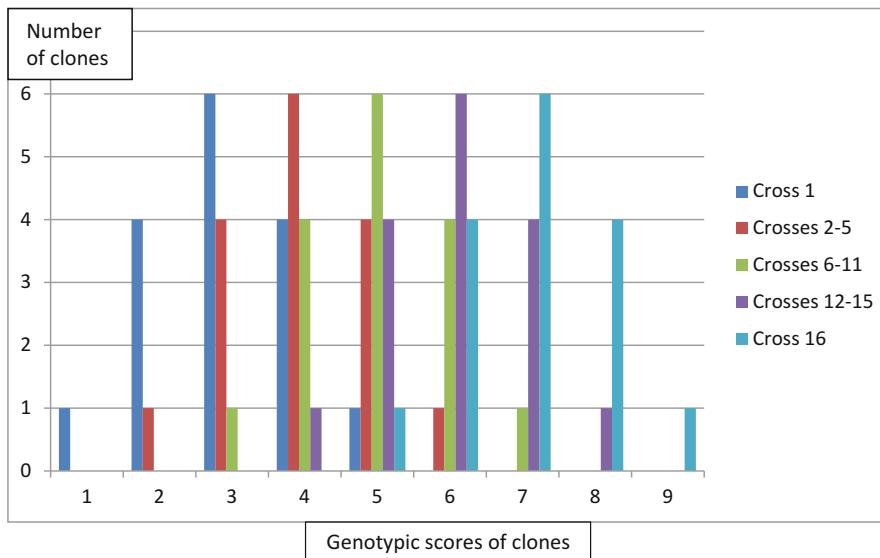
These points can be illustrated with the theoretical example shown in Fig. 11.1 which is based on the binomial distribution for simplicity. Sixteen clones are produced from each of 16 crosses. The means of the 16 clones are 3 for the first cross, 4 for the next four, 5 for the next six, 6 for the next four and 7 for the final cross. The variance of the cross means is 1.0. The variation between clones within a cross is the same for each cross and also has a variance of 1.0. Hence the variation between the crosses is the same as the variation within crosses. If we decide to select clones with scores greater than the mean of the best cross (i.e.,  $>7.0$ ), then this best cross will contribute five clones, the four crosses ranked second equal will each contribute one clone and the other eleven crosses will not contribute at all.

Wricke and Weber (1986) used a different approach to hybridization strategy in their book on selection in plant breeding. They considered maximization of the response to one-step selection for individual clones, using a linear model for the phenotypic value ( $P_{jk}$ ) of genotype (clone)  $k$  from cross  $j$ , in a crossing strategy involving  $m$  crosses each producing progeny of size  $n$  for a total of  $N$  ( $=mn$ ) offspring.

$$P_{jk} = \mu + C_j + S_{jk} + \bar{e}_{jk}$$

where  $\mu$  is the general mean,  $C_j$  is the effect of cross  $j$ ,  $S_{jk}$  is the effect of genotype  $k$  within cross  $j$  and  $\bar{e}_{jk}$  is the environmental variation appropriate for the trial design.

There is an intraclass correlation within a cross (related clones which give information about the cross) which influences the selection response. Given the values of the genetic variation between and within crosses, and the environmental variation, Wricke and Weber (1986) found numerically the optimum values of  $m$  and  $n$  for a given  $N$ . The number of crosses increased with the genetic variance between crosses and decreased with an increase in the environmental variance. Fifty crosses each of size 24 for 1200 offspring was optimum when the three variances were equal and the selected proportion was from 10 to 20 %. The authors acknowledge that their examples are only a rough guide since the exact variances

**a****b**

**Fig. 11.1** Variation between and within 16 crosses where 16 clones have been produced from each cross. (a) Histogram of 16 crosses showing variation between their means. (b) Histogram of 16 clones from each cross with a different mean

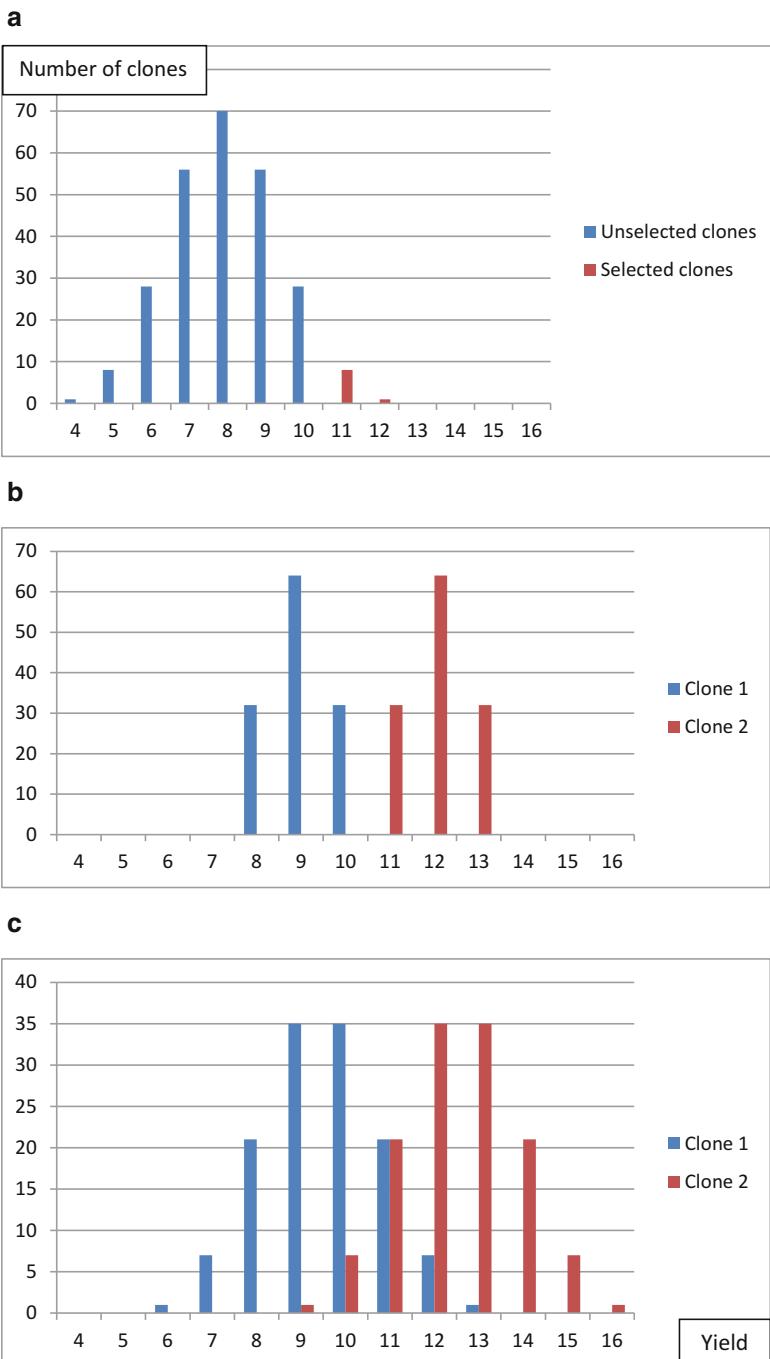
were not known; indeed, are not known until the crosses have been made and assessed. I think that it fair to conclude that the optimum hybridization strategy for a breeder is still a difficult problem.

## Theory of Clonal Selection: Multistage

The theory and practice of clonal selection was explained in a classic paper by Finney (1958) on plant selection. He recognized three stages of selection. In the first stage a breeder (visually) discards seedlings and clones with obvious defects while building up material for more detailed assessment. In the second stage, trials are done to find those clones with the highest merit among those that survive the preliminary screening. Yield assessment is a major component of merit, and Finney concentrated on yield. In the third phase, potential cultivars undergo extensive evaluation in (official) replicated trials over sites and seasons before the best are commercialized if they are better than current cultivars. Finney's paper was concerned with the second stage in which  $N$  clones are reduced to  $n$  over  $k$  clonal generations (years for an annual crop) where  $\pi$  is the proportion selected so that  $n = \pi N$ . The theory also applies to selection between inbred lines or hybrids undergoing trial. He came to a simple tentative conclusion to a complex problem. The optimum strategy for maximizing the gain from selection is never far from the following one. Select an equal proportion ( $P$ ) of clones each generation so that  $P_1 = P_2 = \dots = P_k = \pi^{(1/k)}$ . Furthermore, with finite resources (e.g., land area =  $A$ ), divide them equally between generations so that  $A_1 = A_2 = \dots = A_k = A/k$ . The number of generations of selection is a balance between diminishing gains from each generation being offset by delays in the commercialization of new cultivars. Finney thought that it would be difficult to justify more than five generations of selection and that two or three would usually be adequate. A simple example with three generations would be as follows: 1080 clones are assessed in small plots in an unreplicated trial and the best 360 ( $\frac{1}{3}$ ) are selected; these are assessed in a trial with three replicates and the same size plots (i.e., same area) and the best 120 ( $\frac{1}{3}$ ) are selected; finally these are assessed in a trial with three replicates and plots three times as large (i.e., same area) and the best 40 ( $\frac{1}{3}$ ) are selected. We now need to examine these conclusions in more detail by considering some key points.

### ***Effect of Environmental Variation on Ability to Select Best Clone***

Let us start by considering Fig. 11.2a. For simplicity this shows the yields of 256 clones as a histogram following a binomial distribution, which in practice



**Fig. 11.2** Risks of rejecting a good clone and selecting an average one. (a) Progeny of size 256 clones: all variation is genetic; population mean is 8. (b) Clones with means of 9 (clone 1) and 12 (clone 2) and little ‘error’ variation. (c) Clones with means of 9.5 (clone 1) and 12.5 (clone 2) and much ‘error’ variation

would approximate to a continuous normal distribution. If all of the variation was genetic differences between clones, there would be no problem in selecting the best clone (yield 12 versus population mean of 8) or the best nine clones (mean yield 11.11 versus 8). The selection could be done in a single generation and the gains over the population mean would be the maximum possible and determined by the proportion of clones selected (1 in 256 and 9 in 256, respectively). Once environmental variation is present, we can no longer be certain of selecting the best clones, and this is what happens in practice. In Fig. 11.2b there is little environmental (error) variation and clone 2 with a genotypic mean of 12 can be distinguished with certainty from clone 1 with a mean of 9. However, in Fig. 11.2c there is a lot of environmental variation and a risk that clone 1 is selected and/or clone 2 is rejected when clones with a phenotypic value greater than 10 are selected. Finney thought that a breeder should accept this situation as inevitable and concentrate on maximizing the gain from selection, rather than worry about maximizing the probability that a good clone will be selected and not rejected. The latter, however, is a consideration when it comes to the credibility of an official trials system. Breeders want to see a fair trials system in which a genetically superior cultivar is selected and an inferior one is rejected (Talbot 1997). As the precision of the trial system is reduced, the probability of rejecting a good genotype increases as does the risk of wrongly selecting poor genotypes.

### ***Maximizing the Gain from Selection***

When we considered the response to selection during population improvement, we had to mate the selected individuals to produce the next generation. Selection in a clonal generation is simpler as the response ( $R$ ) is simply the selection differential ( $S$ ) multiplied by the broad sense heritability ( $V_G/V_P = V_G/(V_G + V_E/r)$ ), where  $V_P$  is the phenotypic variation between clones,  $V_G$  is the genetical variation between clones,  $V_E$  is the environmental variation from plot to plot and  $r$  is the number of replicates (plots) of each clone. The selection differential  $S$  is the phenotypic difference between the mean of the selected group of clones ( $M_S$ ) and the mean of the unselected population ( $M$ ). The standardized selection differential  $S/\sigma_P$  is called the intensity of selection  $i$ , where  $\sigma_P$  is the square root of  $V_P$  (and  $\sigma_G$  is the square root of  $V_G$ ).

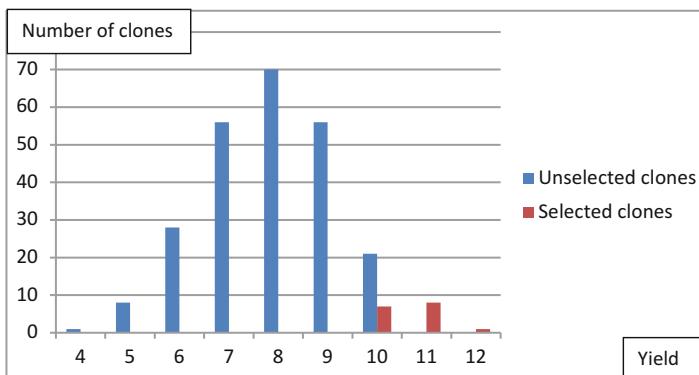
$$R = (V_G/V_P)S = i(\sigma_G/\sigma_P)\sigma_G$$

Hence the response is proportional to the square root of the genetic variance and has its maximum value when the environmental variance is zero ( $\sigma_G = \sigma_P$ ).

### ***Single-Stage and Two-Stage Selection***

Figure 11.3 illustrates single-stage selection with  $V_G = 1$ ,  $V_E = 1$  (unreplicated plots say) and  $V_P = 2$ ; and hence a heritability of 0.5. If the best 16 clones are selected, their phenotypic mean is 10.625 and the selection differential is 2.625. The predicted response to selection is therefore 1.3125. The observed response in the next clonal generation can be calculated from the table in Fig. 11.3. The mean genotypic value of the 16 selected clones is indeed 9.3125, giving a response to selection of 1.3125.

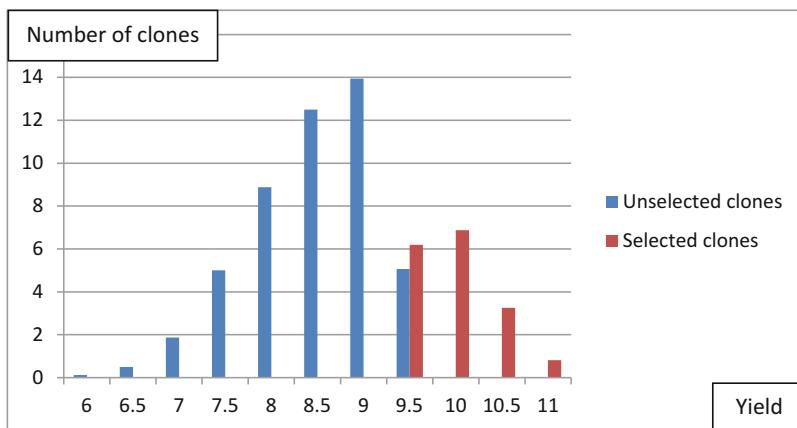
Now we need to consider if we can do better through two-stage selection. This time let us select the best 65 clones in the first stage, rather than 16. They have a genotypic mean of 8.8615. We now assess them in a replicated trial so that the

**a****b**

				<i>G</i>			
	Yield	Frequency	6	7	8	9	10
	4	1	1	0	0	0	0
	5	8	4	4	0	0	0
	6	28	6	16	6	0	0
	7	56	4	24	24	4	0
<i>P</i>	8	70	1	16	36	16	1
	9	56	0	4	24	24	4
	10	28	0	0	6	16	6
	11	8	0	0	0	4	4
	12	1	0	0	0	0	1
Select	16	0	0	1.5	8	6.5	
Select	65	0	2	18	32	13	

**Fig. 11.3** Single-stage selection: progeny of size 256 clones, genetic variation (values 6, 7, 8, 9 and 10 in a binomial distribution: ratios 1:4:6:4:1) and environmental variation (values -2, -1, 0, 1, 2 in a binomial distribution: ratios 1:4:6:4:1); population mean is 8. (a) Yields of 256 clones. (b) Phenotypes (*P*) and Genotypes (*G*) of the 256 clones

clonal means are assessed with greater precision. Following Finney's advice the trial has four replicates so that the 65 clones occupy about the same area of land as the 256 clones in the previous generation. The environmental component of the phenotypic variation in clonal means is therefore one quarter of the plot to plot variation. The genetic variance is also less as a result of selecting 65 out of 256 clones. The outcome is  $V_G = 0.581$  (reduced from 1),  $V_E = 0.250$  (reduced from 1),  $V_P = 0.831$  and a heritability of 0.699, which is a modest increase on 0.50 (Fig. 11.4). If we now select the best 16 clones (the ones with score 9.5 being selected at random for the purpose of this exercise), we find that they have a

**a****b**

				G		
	Yield	Frequency	7	8	9	10
	6	0.125	0.125	0	0	0
	6.5	0.5	0.5	0	0	0
	7	1.875	0.75	1.125	0	0
	7.5	5	0.5	4.5	0	0
P	8	8.875	0.125	6.75	2	0
	8.5	12.5	0	4.5	8	0
	9	13.9375	0	1.125	12	0.8125
	9.5	11.25	0	0	8	3.25
	10	6.875	0	0	2	4.875
	10.5	3.25	0	0	0	3.25
	11	0.8125	0	0	0	0.8125
	Select	16	0	0	5.6	10.4

**Fig. 11.4** Two-stage selection: progeny of size 65 clones from stage one, genetic variation (values 7, 8, 9 and 10 for 2, 18, 32 and 13 clones) and environmental variation (values  $-1, -\frac{1}{2}, 0, \frac{1}{2}, 1$  in a binomial distribution: ratios 1:4:6:4:1); population mean is 8.8615. (a) Yields of 65 clones. (b) Phenotypes ( $P$ ) and Genotypes ( $G$ ) of the 65 clones

phenotypic mean of 9.9941 and a genotypic mean of 9.65. Hence  $S = 1.1326$ ,  $R = 0.7885$ , and  $R/S = 0.696$  (the realized heritability), which is very close to the estimated heritability. The genotypic mean of 9.65 is higher than the 9.3125 achieved for 16 clones from one-stage selection. This modest advantage is achieved at the expense of taking 2 years instead of one to do the selection.

Finney (1958) concluded that most of the possible gain can be achieved in two (rather than more) stages with moderate experimental error and plant selection, and that two-stage selection is little better than one-stage selection under conditions of very high experimental error. Finney (1958) also concluded that when the environmental variation is large compared with the genetic variation, only a small gain can be achieved in one-stage selection. Nevertheless, it is worth discarding a large proportion of clones at random before assessment starts to reduce the experimental error (larger plots or more replicates) for those clones that do go into the yield trial, even though intense selection is not now possible.

### ***Other Approaches***

Wricke and Weber (1986) compared Finney's approach to that of Cochran (1951) who regarded selection in two stages as selection for two traits using independent culling levels (see next section). They concluded that the selection response using Finney's values is only slightly less than the response using the optimum values for the selection fractions and number of replicates calculated by Cochran's method. Furthermore, little improvement comes from using information from the first stage in the second stage. Hence plant breeders can still benefit from reading the whole of Finney's paper, despite the fact it was written over 50 years ago.

A more recent paper by Ishii and Yonezawa (2006) has shown how the work of Finney and other researchers can be extended to provide more general guidelines for multi-stage yield selection. For many crops their paper is more relevant to Finney's third stage of plant breeding where potential cultivars undergo extensive evaluation in replicated trials over sites and seasons before the best are commercialized if they are better than current cultivars. The authors assumed that a breeder would not proceed to stage two if none of the entries in the first stage was superior to a standard check. They used extensive Monte Carlo calculations (simulations), based on the principle of maximizing the probability of success ( $S$ ) that an entry with the highest yielding ability is selected for a target region from a population of  $N$  candidates. They considered  $N$  in the range 20–100, with examples given for 50, and three-stage yield selection. The entries were assessed in replicated trials at a number of locations in the target region. The results confirmed that the total number and allocation of locations, not of replications, determined  $S$ . With the same total number of locations ( $s_1 + s_2 + s_3$ ), a uniform allocation over stages ( $s_1 = s_2 = s_3$ ), or decreasing number ( $s_1 > s_2 > s_3$ ) in the presence of a high magnitude of genotype  $\times$  location interaction, was nearly if not exactly the best. If circumstances permit, new locations from the target region should be chosen in each stage.

The common allocation practice of  $1 (=s_1) < s_2 < s_3$  proved undesirable, and hence should only be used when the amount of planting material is the limiting factor. The optimum selected fractions ( $p = 1/N = p_1 p_2 p_3$ ) and proportions of total plots ( $\alpha_1 + \alpha_2 + \alpha_3 = 1$ ; where total was predetermined as  $N\beta$  with  $\beta$  in the range 2–10) used at each stage were in the order  $p_1 < p_2 < p_3$  and  $\alpha_1 > \alpha_2 > \alpha_3$ . However,  $p_1 > p_2 > p_3$  instead of  $p_1 < p_2 < p_3$  is appropriate when genotype  $\times$  year interaction is the major error component in the yield scores of entries. Replications should be allocated once all of the other decisions have been made.

## Theory of Clonal Selection: Multitrait

So far in this chapter we have talked about desired traits, an index of merit, and yield as a major component of merit. In a practical breeding programme the breeder will most likely want to improve yield, quality traits and resistances to the most serious abiotic and biotic stresses. In the last chapter three strategies were mentioned for such multitrait selection: tandem selection, independent culling levels and index selection. In tandem selection, only one character is selected in each cycle, say trait one in the first cycle, trait two in the second, and so on. With independent culling levels, all clones are discarded which have a phenotypic value below the culling level for at least one trait, e.g., low yield, poor quality or poor disease resistance. With index selection, the aim is to improve traits simultaneously in such a way as to make the biggest possible improvement in overall merit.

Bulmer (1980) gave the following simple explanation of why index selection is superior to selection by independent culling, which in turn is superior to tandem selection, as first demonstrated by Hazel and Lush (1942). Let us suppose that  $k$  traits are independently and normally distributed with the same variance ( $V$ ) and the same heritability ( $h^2$ ). If we select a proportion ( $p$ ) from a large population, the selection intensity ( $i$ ) for index or tandem selection is  $Z/p$ , where  $Z$  is the standard normal ordinate corresponding to  $p$ . The selection intensity can therefore be determined from tables of the properties of the normal distribution. Under independent culling we set the threshold for each trait in such a way that the probability of its being exceeded is  $p^* = p^{1/k}$ . This ensures that the proportion selected is  $p$ . The selection intensity ( $i$ ) for any trait considered by itself is  $Z^*/p^*$ , where  $Z^*$  is the standard normal ordinate corresponding to  $p^*$ . The responses to selection in this simple example can be calculated from the formula for a single trait:

$$R = h^2 S = h^2 i \sigma_P = h^2 (Z^* / p^*) \sigma_P$$

where  $S$  is the selection differential and  $\sigma_P = V^{1/2}$ .

With tandem selection, only one trait is selected each generation, so the genetic gain per generation is:

$$R = h^2(Z/p)V^{1/2}.$$

With independent culling levels,  $k$  traits are selected. For each trait  $R = h^2(Z^*/p^*)V^{1/2}$ , and assuming each trait makes the same contribution to overall merit, on the scale of overall merit, the total response is simply the response for each trait multiplied by the number of traits:

$$R = kh^2(Z^*/p^*)V^{1/2}.$$

With index selection, the contributions to overall merit from each trait for each clone are first determined, and then selection practised on the merit score. As the  $k$  traits are independently and normally distributed with the same variance ( $V$ ), the variance of the merit scores is  $kV$ . Hence the response to selection for merit is:

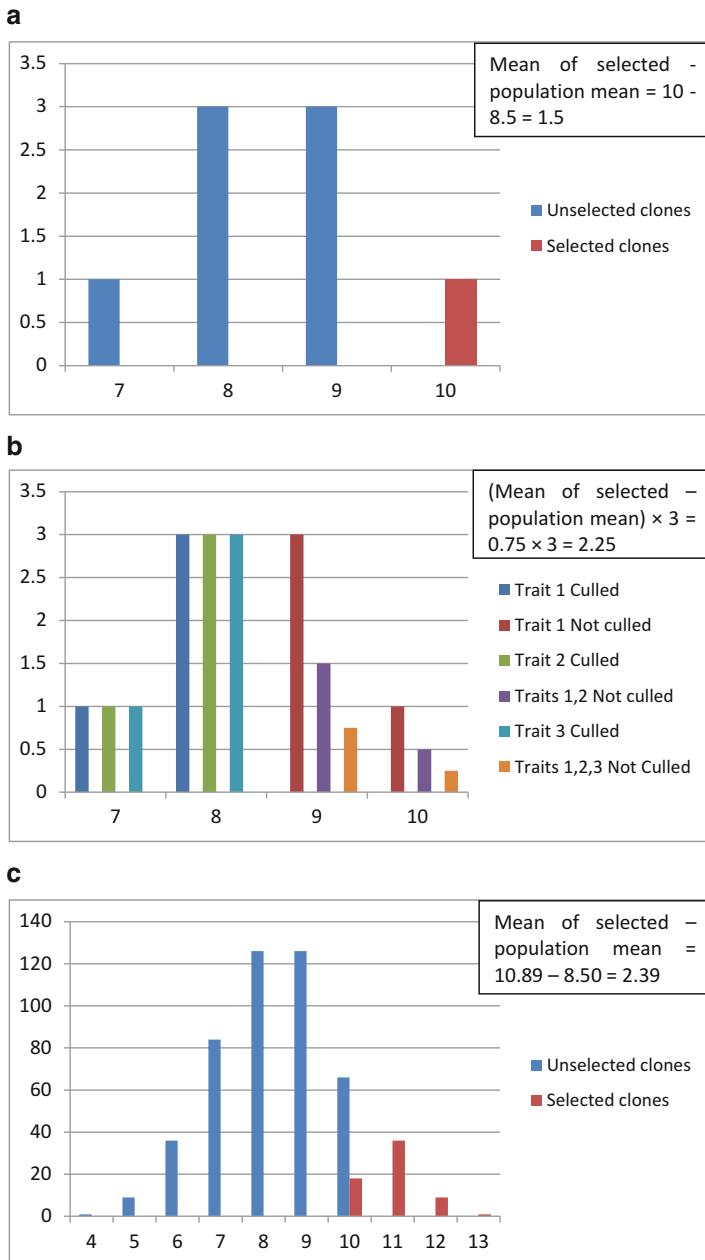
$$R = h^2(Z/p)(kV)^{1/2}.$$

The relative efficiency of tandem selection to index selection is  $h^2(Z/p)V^{1/2}$  divided by  $h^2(Z/p)(kV)^{1/2}$ , which equals  $1/k^{1/2}$ . With  $k=4$ , for example, the relative efficiency is 0.5.

The relative efficiency of independent culling levels to index selection is likewise  $k^{1/2}(Z^*/p^*)(p/Z)$ . With  $k=4$  and  $p=0.1$ , the relative efficiency is 0.8 ( $2 \times 0.697 \times 0.570$ ).

For all  $k$  and  $p$  the genetic gain is in the rank order Index selection > Independent culling > Tandem selection. This conclusion can be extended to the more general case when the traits are correlated and of varying economic value and variability. Young (1961) found that the superiority of the index over other methods increases with an increasing number of traits under selection, but decreases with increasing differences in relative importance, its superiority being at a maximum when the traits are of equal importance. The superiority of the index over independent culling levels decreases with increasing selection intensity, but its superiority over tandem selection is independent of intensity. The relative efficiency of the index over other methods is much affected by the phenotypic correlation between traits when the traits are of equal importance, the relative efficiency of the index being higher when the phenotypic correlation is low or negative.

The differences between the three types of selection are shown graphically in Fig. 11.5, using the binomial rather than the normal distribution for simplicity. In tandem selection, one eighth of the clones are selected for one trait with a mean of 8.5 and variance of 0.75. The response to selection is 1.5. With independent culling levels, one half of the clones are culled independently for each of three traits so that one in eight clones is again selected. Of these, on average, three quarters score 9 and one quarter scores 10 for each trait so that selection of each trait contributes 0.75 to merit and hence an overall response to selection of 2.25. With index selection, three traits with the same variances for merit as those for tandem selection and independent culling are combined into an index (mean 8.5, variance  $3 \times 0.75$ ) and then



**Fig. 11.5** Multitrait selection. (a) Tandem selection: 1 out of 8 clones selected for one trait. (b) Independent culling levels: 4 out of 8 clones culled independently for each of three traits leaving 1 out of 8 clones ( $\frac{1}{4}$  score 9 and  $\frac{1}{4}$  score 10 on average for each trait). (c) Index selection: 64 out of 512 (1 out of 8) clones selected on index of three traits

64 out of 512 (1 out of 8) clones are selected on the index. Eighteen out of the 84 scoring 10 are chosen at random to achieve the total of 64 selected. The overall response to selection is 2.39. The order, as expected, is Index selection > Independent culling > Tandem selection.

## Practice of Multitrait Selection

It is usually not possible to assess all traits in all clonal generations, particularly the early generations when there is a lack of planting material and the assessment plots may not be typical of agricultural practice. This is particularly true of the seedling generation, and to a lesser extent, the first few clonal generations. Inevitably there is some tandem selection in which one trait is selected in one generation, another trait in the next generation and so on, despite the fact that it is not an efficient procedure. This is because once material has been discarded on one trait, it cannot be assessed for other traits. Once a number of traits can be assessed in a generation, the breeder needs a strategy for discarding clones that will not continue into the next generation. In the previous chapter index selection was dealt with as the most efficient way of selecting parents for the next round of crossing: excellence in some traits is allowed to outweigh average performance in other traits or even poor performance in a trait. However, independent culling levels is the method most commonly used to select clones for further assessment as potential cultivars, and the same is true for the selfing generations with inbreeding crops. With independent culling levels, all clones are discarded which have a phenotypic value below the culling level for at least one trait, e.g., low yield, poor quality or poor disease resistance. The appeal of the method is its simplicity and the fact that culling levels can be set to achieve improvements over currently popular cultivars (industry standards). The problem is that none of the clones may meet the criteria and the breeder is then faced with deciding the relative importance of the different traits; i.e., where in effect to relax the culling levels. Independent culling levels are never in fact as good as an optimum index because clones which are outstanding for many traits can be discarded for not meeting the culling point for just one trait. Furthermore, traits can differ in their heritabilities and correlations can occur between traits. Desirable ones will result in more clones being selected whereas undesirable ones will result in fewer. The general theory of the response to selection with independent culling levels is difficult (Wricke and Weber 1986), but some useful points can be made from considering uncorrelated traits. Furthermore, as breeders accumulate relevant data, they can simulate and compare the results of different selection strategies. Breeders should certainly always monitor progress against appropriate control cultivars throughout their programmes so that any lack of expected progress can be detected and investigated.

## Uncorrelated Traits

When traits are uncorrelated, the response to selection for each trait is simply  $R = h^2 S = ih^2 \sigma_P$ , where  $i$  is the intensity of selection,  $h^2$  is the broad sense heritability and  $\sigma_P$  is the phenotypic standard deviation for that trait.

Let us consider the simple example in Table 11.1 which shows the bivariate distribution of phenotypes for 256 clones assessed for two uncorrelated traits. Now let us cull all those clones with a phenotypic score of 7 or less for trait 1 and all of those clones with a phenotypic score of 9 or less for trait 2. The result is 11 clones out of 256 remain, or 4.3 %. In other words, 68.75 % of clones are retained after culling for trait 1, and  $68.75\% \times 6.25\%$  are retained after culling for both traits. The mean of the 11 clones is 8.55 ( $S = 0.55$ ) for trait 1 and 10.00 ( $S = 2$ ) for trait 2. If the heritabilities of the two traits were 0.1 and 0.9, the responses to selection ( $R$ ) would be 0.055 ( $0.1 \times 0.55$ ) and 1.8 ( $0.9 \times 2$ ), respectively. The means of the 11 selected clones in the next clonal generation would be 8.055 for trait 1 and 9.8 for trait 2. However, if the heritabilities were the other way round, namely 0.9 and 0.1, the responses to selection ( $R$ ) would be 0.495 ( $0.9 \times 0.55$ ) and 0.2 ( $0.1 \times 2$ ), respectively. The means of the 11 selected clones in the next clonal generation would be 8.495 for trait 1 but only 8.2 for trait 2. Relative to the means of the two traits, the breeder might conclude that responses of 0.055 and 1.8 are collectively better than responses of 0.495 and 0.2, and that the right decision is to apply the greatest selection intensity to the trait with the highest heritability. The breeder might also conclude that there are optimum selection intensities to achieve the best compromise for the two traits. However, the culling levels were set relative to standards for a good reason so it is responses relative to the standards that are important. What is clear is that the culling level of 9 or less for trait 2 achieves a mean of only 8.2 when the heritability is very low at 0.1. We are back to the point made earlier. As the precision of a trial system is reduced, the probability of rejecting a good clone increases as does the risk of wrongly selecting poor clones. The use of independent culling levels does not solve the problem. In practice, a breeder usually delays selection of traits subject to a lot of environmental variation until replication, blocking and spatial analysis can reduce its effects.

Having considered the theory of clonal selection, it is now time to see what breeders have actually done in a number of clonally propagated crops.

**Table 11.1** Phenotypic values of 256 clones for two uncorrelated traits with same means (8) and variances: Trait 1  $\leq 7$  culled, Trait 2  $\leq 9$  culled leaving clones shown in bold

		Trait 2					
		Phenotypes	6	7	8	9	10
Trait 1	10		1	4	6	4	<b>1</b>
	9		4	16	24	16	<b>4</b>
	8		6	24	36	24	<b>6</b>
	7		4	16	24	16	4
	6		1	4	6	4	1

## Root and Tuber Crops

We are going to consider the potato (*Solanum tuberosum*), cassava (*Manihot esculenta*) and sweet potato (*Ipomoea batatas*). The potato is a perennial herb that is grown as an annual crop for its tubers. Cassava is a perennial, woody shrub that produces large fleshy storage roots within 1–3 years after planting. Sweet potato is a trailing herbaceous perennial that is grown as an annual crop for its storage roots. The potato was domesticated in South America, cassava probably in South America but possibly in Mexico, and sweet potato probably in Mexico but possibly in South America, all some 8000 years ago (Bradshaw 2010). Much later, after Columbus discovered the New World in 1492, European sailors introduced the potato to Europe and from there to many other parts of the world, and both cassava and sweet potato to Africa and then Asia. Interestingly, the sweet potato was being grown in Oceania before Columbus, but the routes of introduction are still debated. Today the potato is the fourth most important food crop in the world after maize, rice and wheat, and the five largest potato producing countries are China, India, the Russian Federation, Ukraine and the USA. Cassava is the most important root and tuber crop in the tropics where it is a primary staple food in many of the poorest countries, with the largest production in Nigeria, Brazil, Thailand and Indonesia. The sweet potato is also a food staple in Asia, Africa, and America, but with production dominated by China, where half of the crop goes for animal feed. The edible storage organs are underground tubers for potatoes and storage roots for cassava and sweet potato. Both roots and tubers store energy as starch, but are also valuable sources of minerals, vitamins and other antioxidants. Whereas potatoes are also propagated through their tubers, sweet potato is normally propagated as vine cuttings, and cassava is propagated as stem cuttings because its roots cannot be used for reproductive purposes. These propagation methods are slow, so more rapid ones have been, and are being, developed for the breeding and multiplication of new cultivars.

The breeding of new cultivars involves sexual hybridization and hence a knowledge of flowering, pollination and seed set. Natural cross-pollination of all three crops is by insects. Self-pollination also occurs in potatoes as gametophytic self-incompatibility breaks down in polyploids, but the crop is not tolerant of inbreeding. Out-crossing is encouraged in cassava by separate male and female flowers and in sweet potato by sporophytic self-incompatibility. Cassava is a functional diploid species ( $2n=2x=36$ ), the principal cultivated potato is an autotetraploid ( $2n=4x=48$ ) and sweet potato is a hexaploid ( $2n=6x=90$ ), probably an allo-autopolyploid as a result of being a hybrid between a diploid and tetraploid species. Since domestication, all of the crops have been improved by both conscious and unconscious farmer selection. More modern hybridization and selection by farmers, hobby breeders and seedsmen occurred for the potato during the nineteenth century, and scientific breeding methods were developed from the early twentieth century onwards. Modern breeding of cassava and sweet potato started in the 1920s, but intensified really only from the 1960s and 1970s. This modern breeding work has been helped by the establishment of International Research Centres aimed at

providing food security and eradicating poverty in developing countries; and this will remain important during a period of human population growth and climate change. As mentioned in Chap. 2, both cassava and potato respond well to increased carbon dioxide concentrations. Genetic transformation should be of particular value in potato and sweet potato as these are vegetatively propagated polyploids with complex inheritance patterns. In this chapter, however, we will focus on conventional breeding, and for potatoes on the vegetatively propagated crop. There are circumstances where potato cultivars based on current methods of true potato seed (TPS) propagation are an attractive proposition, as explained by Bradshaw and Bonierbale (2010).

## Potato Breeding

### *Modern Potato Breeding*

The first named potato cultivars in Europe and North America can be traced to the 1730s (Reader 2008), and this marked the start of the transition from landraces to modern cultivars. In other words, distinct clones were being selected, multiplied and marketed from seedlings raised from seeds in the berries from natural open-pollination. The parentage of the resulting cultivars was usually recorded as unknown. Thomas Andrew Knight (Knight 1807) advocated the use of artificial hybridization in the breeding of new potato cultivars, but James Clark in England was one of the first breeders to produce successful cultivars from such deliberate cross-pollinations, for example, Magnum Bonum in 1876 (Wilson 1993). Potato breeding certainly flourished in Europe and North America during the second half of the nineteenth century when exchanges of germplasm started to occur, and many new cultivars were produced by farmers, hobby breeders, and seedsmen. Even then, the raising of seedlings from seed of self-set berries remained a common practice which continued into the twentieth century. North America's most popular potato cultivar, Russet Burbank, released in 1914, was descended from Rough Purple Chili through three generations of open-pollination (Ortiz 2001). Modern potato breeding in India and China started later in the 1930s, but with rapid expansion since 1948 and 1978, respectively (Gaur and Pandey 2000; Jin et al. 2004). The extent of progress since 1807 can be judged by the latest World Catalogue of Potato Varieties (Pieterse and Hils 2009) which lists more than 4500 cultivars from 102 countries covering all potato growing regions in the world.

### *Scientific Potato Breeding in Scotland*

Scientific potato breeding developed from the beginning of the twentieth century and the programme at the Scottish Plant Breeding Station (SPBS) from 1920 and the Scottish Crop Research Institute (SCRI) from 1980 was typical. A complete

record exists of all the crosses made from 1920 onwards, and this provides a genetic history which has been analysed by Bradshaw (2009a). Seventy-two cultivars had been bred by 2008, and all came from pair crosses between clones and cultivars that complemented each other for desirable characteristics. The methods used for deliberate artificial hybridizations are briefly described by Bradshaw and Bonierbale (2010) with references to more detailed accounts. A temperature of 19 °C and 16 h of daylength are recommended for crosses involving *S. tuberosum* Group Tuberosum, whereas Group Andigena is in effect day-neutral for flowering, but not for tuberization. Breeders usually encourage flowering in a glasshouse by the periodic removal of daughter tubers, but not all desired crosses are successful. Problems can arise from clones failing to flower, buds and flowers dropping either before or after fertilization, low pollen production, and failure to produce viable pollen (male sterility). The parentages of cultivars used in the early breeding work trace back either to old European/North American cultivars or to cv. Rough Purple Chili, with the exception of cv. Dr McIntosh which has another cultivated species in its pedigree, *Solanum rybini* (= *S. phureja* or *S. tuberosum* Group Phureja).

Some of the cultivars and clones subsequently used as parents were the result of introgression of desirable genes from the wild and cultivated species of Latin America, maintained in the Commonwealth Potato Collection (CPC) since 1939 (Bradshaw and Ramsay 2005). Introgression of late blight (*Phytophthora infestans*) resistance began in 1932, followed by virus resistance from 1941, and finally potato cyst nematode (*Globodera rostochiensis* and *G. pallida*) resistance from 1952. Just seven of the 219 wild tuber-bearing species recognized by Hawkes (1990) feature in the pedigrees of the cultivars, with *Solanum demissum* for late blight resistance in 58, *S. vernei* for nematode resistance in 19 and *S. microdontum* for *Potato virus Y* resistance in 15. Fifteen out of the 50 cultivars starting with Pentland Javelin in 1968, have the *H1* gene for resistance to *G. rostochiensis* introgressed from group Andigena potatoes (the short-day-adapted landraces of the Andes). In contrast, just one cultivar (Shelagh) has a parent from a long-day-adapted Andigena (Neotuberosum) population, derived from accessions in the CPC by recurrent mass selection. In other words, broadening the genetic base of potato breeding from landraces has not had much impact compared with introgression.

Consumer quality was always a key objective for achieving commercial success with new cultivars, and only cultivars with acceptable quality traits have been released. Up until 1970 new cultivars were assessed primarily for cooking and keeping quality, although cv. Pentland Dell (National Listed in 1960) went on to be primarily used for processing into French fries, and was still the twelfth most widely grown cultivar in Britain in 2013. From 1970 onwards, new cultivars were assessed for consumer quality, which now included processing quality as well as cooking quality for table use. The period since 1960 in Britain had seen a steady increase in the proportion of the crop which is processed, to about 50 % in 2008. The main change in the production and sale of fresh potatoes is that today 63 % are supplied by just four supermarket chains, having been packed by a similar small number of companies with their dedicated seed producers and ware growers.

In the early days the programme was small in size, both in terms of number of crosses made and number of seedlings raised. As a consequence, for crosses made up until 1960, the time from crossing to Registration or National Listing of a new cultivar was only 8.5 years on average. Subsequently the programme increased in size, more traits were assessed, and potential cultivars were trialled in more environments. As a consequence, the average time to National Listing increased to 13 years. By 1981, advances in computing hardware and the in-house development of suitable software allowed randomization and replication of all trials of all clones undergoing selection, as well as facilitation of overall planning and data management. The target length of time with the breeding scheme was 12 years, and this was achieved with a number of cultivars. Furthermore, the scheme was typical of most relatively large programmes both then and now.

### ***Breeding Scheme***

The seedling generation (SG) in the glasshouse comprised 100,000 genetically unique seedlings from some 200–300 crosses made the previous year. Visual selection reduced this number of potential cultivars to 1000 clones in replicated yield trials at a ware site in the third clonal generation (TCG). The first clonal generation (FCG) comprised 50,000 spaced plants at a high-grade seed site with a short growing season, rather atypical of normal ware production. The second clonal generation (SCG) of 4000 unreplicated three- or four-plant plots was also grown at the seed site. Decreasing numbers (1000, 500, 200) of potential cultivars were then assessed for 3 years in replicated yield trials (with plots of 5–10 tubers) at a local ware site before the most promising ones (60, 10, 5) underwent 3 years of more extensive testing at a number of sites, including overseas ones. One or a few clones would then be entered into 2 years of official statutory trials (National List Trials) and registered for Plant Breeders' Rights. Multiplication from disease-free stock would start with a view to commercialization. During these intermediate and final stages of selection the production of seed tubers was separated from the trials that were grown under ware conditions, designed as far as possible to resemble those of good commercial practice. Clones undergoing selection were assessed for their yield and agronomic performance, external and internal defects, cooking and processing characteristics, and for their disease and pest resistance in special tests. More detailed information on handling the intermediate and later generations, including the conduct of yield trials and selection for disease and pest resistance, can be found in the review by Bradshaw (2007). The selection criteria and testing procedures in such a multitrait, multistage scheme have largely been governed by practical considerations and experience of the reliability of the various tests used, rather than by genetical knowledge, such as heritabilities and genetic correlations between traits. They have also been governed by the low tuber multiplication rate each generation (often around times 8 on average). The theory for maximizing the response to clonal selection with finite resources is in fact complicated because the

genotypic variance between clones changes with each generation of selection, as explained at the beginning of this chapter (Wricke and Weber 1986; Bradshaw and Mackay 1994a, b).

### ***Analysis of Breeding Scheme: Crossing Strategy***

The breeding scheme just described was geared to the many economically important traits that were viewed as complex polygenic traits. Many traits are still best viewed this way and the breeder still has to rely primarily on phenotypic data and the concepts of quantitative genetics to determine crossing strategy and the population size of progenies. With a highly heritable trait like fry colour the midparent value is a good predictor of the mean performance of the offspring and a few carefully chosen crosses with larger progeny sizes can be made where processing quality is the main target. However, progeny sizes in excess of 1000 may not be justified as the greater extremes may be difficult and expensive to find. In contrast, with only a moderately heritable trait, such as yield, offspring mean is less predictable and more crosses need to be made to ensure that they include the best possible ones.

Other things being equal, the highest yielding offspring clone from a set of crosses will come from transgressive segregation (offspring superior to the better parent) in the highest yielding cross, the latter resulting from parents with high general combining abilities (GCAs) and a positive specific combining ability (SCA). Bradshaw (2009b) gave an example of these sources of variation from a  $6 \times 6$  half diallel set of 15 crosses, in which 218 offspring clones (harmonic mean of 14.31 clones per cross) were assessed at SCRI in Dundee in 1993 and 1994. There was more variation within (genetic variance 3.168) than between crosses (0.954), probably a reflection of high yielding parents being chosen for the investigation, as well as smaller interactions with years. SCA (0.457) accounted for slightly more variation than GCA (0.354), and the interactions with years were smaller. There was a high correlation between parental yields and their GCAs ( $r = 0.875$ ) and a moderate one between offspring means and mid-parent values ( $r = 0.69$ ). The broad-sense heritability of the 218 offspring clones was 0.785, where clone means were averaged over two replicates in each of 2 years and clone  $\times$  year interactions were included in the phenotypic variation. Just 17 out of the 218 (8 %) offspring clones out yielded the best parent ‘Cara’, with the two highest yielding clones coming from the crosses ‘Cara’  $\times$  ‘P. Squire’ (14.30 kg/plot) and ‘Desiree’  $\times$  ‘M. Piper’ (13.73). The components of these means were as follows:

$$14.30 (\text{best clone}) = 8.57 (\text{mean of all clones}) + 0.69 (GCA_{\text{Cara}}) + 0.34 (GCA_{\text{P.Squire}})$$

$$+ 0.56 (\text{SCA}) + 4.14 (\text{clone in cross})$$

$$13.73 (\text{second best}) = 8.57 (\text{mean of all clones}) + 0.22 (GCA_{\text{Desiree}})$$

$$+ 0.66 (GCA_{\text{M.Piper}}) + 1.54 (\text{SCA}) + 2.74 (\text{clone in cross})$$

These actual yields should be multiplied by the heritability to give the Best Linear Unbiased Predicted yields. In both crosses, there is transgressive segregation and both parents have positive GCAs and combine with a positive SCA. However, none of the 17 high yielding offspring were good enough for other traits to be selected as a new cultivar. This emphasizes the need for accumulated progress over a number of generations by efficient genotypic recurrent selection. The response to selection can be predicted from the component of variance for GCA and the narrow-sense heritability (Bradshaw 1994b).

### ***Analysis of Breeding Scheme: Early Generation Selection***

A seedling generation of size 100,000 from 200 to 300 crosses would appear reasonable. However, research in the 1980s found that intense early-generation visual selection for most quantitative traits was very ineffective, particularly between seedlings in a glasshouse and spaced plants at a seed site (Bradshaw and Mackay 1994b; Bradshaw et al. 1998). Selection for tuber skin and flesh colour and shape can however be done if these are important for particular consumers. The solution to ineffective early-generation visual selection developed and implemented at SCRI from 1985 was the use of progeny tests to discard whole progenies (= full-sib families) before starting conventional within-progeny selection at the unreplicated small-plot stage (Bradshaw and Mackay 1994b; Bradshaw 2009a). Seedling progeny tests were developed and used for resistance to late blight (both foliage and tuber), resistance to the white potato cyst nematode (*G. pallida*) and tuber yield and appearance, as visually assessed by breeders. Tuber progeny tests were developed and used for fry colour and a second visual assessment of tuber yield and appearance. More true seed of the best progenies would be sown to increase the number of clones on which to practise selection in seeking new cultivars. Examples can be found in the review by Bradshaw (2009a). The conclusion from these breeding programmes was that 40 crosses are adequate for a particular target, with approximately 2000 seedlings from all crosses raised in the non-destructive seedling and tuber progeny tests, and a further 2000 seedlings subsequently raised from the best ten progenies. Thus, in total, 20,000 rather than 100,000 seedlings would be raised from 200 crosses, and in this respect the SPBS/SCRI breeding programmes came to differ from traditional potato breeding.

Continued progress depends upon making further rounds of crosses followed by clonal selection; in other words, on efficient recurrent selection. In the SPBS/SCRI programmes it was found that promising clones could be identified for use as parents after both the first and the second year of ware (yield) trials. Hence combined between and within full-sib family (cross) selection can operate on a 5- or 6-year cycle to keep the momentum of the programme going (Fig. 11.6).



**Fig. 11.6** SCRI potato breeding programme 1991–2009: progeny tests used to discard whole progenies in years 2 and 3, followed by clonal selection within best progenies in year 4 (unreplicated small plots at seed site) and in year 5 (yield trials at ware site), followed by further clonal selection and next round of crossing and selecting (PCN = potato cyst nematode, DM = dry matter content) (Source: SCRI)

This could be reduced to a 3-year cycle with only limited within-family selection; something a breeder might wish to do at the start of a new programme to combine genes from more than two parents as quickly as possible. Such a recurrent selection programme can accommodate new breeding objectives and germplasm whilst continuing to make progress, as demonstrated by Bradshaw et al. (2009b). The temptation to argue that more information is required on a clone before it can be used as a parent should be avoided. If one waits until the clone is entered into, or completes, National List Trials, one could be looking at a 10- or 12-year cycle. Furthermore, successful potato cultivars can have a long life (e.g., Maris Piper has been the number one cultivar in Britain for over 30 years), and it is also tempting to continue to use them as parents, a practice which must be hindering progress.

## ***Genetic Modification, Marker-Assisted Selection, Genomic Selection and Phenotypic Selection***

The current challenge for potato breeders is to integrate marker-assisted selection and genomic selection with phenotypic selection, and also to decide when a transgenic approach is more appropriate (see Chap. 17), given consumer acceptability of the latter. As diagnostic molecular markers become available for major genes and QTL alleles of large effect, potato breeders will be able to select for these desirable genes in greenhouse grown plants, having made sure that they are present in the parents of the cross. However, as QTL alleles of large effect don't explain all of the genetic variation for a trait, subsequent phenotypic selection will be required to achieve the greatest improvement in the trait. Where a high proportion of the genetic variation for a trait is explained by a number of QTL alleles each of small effect, the breeder needs to decide between marker-assisted selection for these alleles, more general genomic selection, and phenotypic selection. The appeal of using molecular markers for marker-assisted or genomic selection is that the selection can be done on greenhouse grown plants, thus saving time and money by not having to do a number of generations of expensive field trials, quality testing and disease testing; and allowing a short cycle time in what is in effect a recurrent selection programme. As the QTLs are unlikely (at present) to explain all of the genetic variation and as genomic selection is likely to be imperfect (at present), phenotypic assessment and selection will be required at some stage. The resultant cultivars, like present day ones, are likely to have a sufficiently long life to make further targeted improvements by genetic transformation worthwhile. The genetically modified cultivar will contain just a few transgenes that have been randomly inserted into its genome. Variation from different random insertions combined with the possibility of somaclonal variation arising during tissue culture, means that the best transformed clone will be sought by phenotypic selection from several hundred produced. In summary, designing a breeding programme is always a challenge.

## ***Seed Potato Production***

Potato yields and quality are certainly best when crops are planted with disease-free seed tubers in the correct physiological state. Experience around the world during the twentieth century showed that this is most likely to be achieved through statutory seed certification schemes operating in areas where potatoes are grown only for seed. Such areas will usually be geographically and climatically less favourable to the aphid vectors of viruses which cause systemic infection (Jeffries et al. 2006). Here we are going to consider asexual reproduction which allows a genetically unique seedling to be maintained, multiplied and grown as a new cultivar. A useful account of the various multiplication procedures can be found

in a special issue of *Potato Research, The Canon of Potato Science* (Struik et al. 2007). A typical seed production scheme with certification is now described.

Seed production starts from pathogen-free microplants which are produced by the Certifying Authority (or under licence) in sterile laboratory conditions. In Scotland, for example, SASA (<http://www.sasa.gov.uk>) is the Certifying Authority for the Seed Potato Classification Scheme (SPCS). Single-node cuttings from tuber sprouts, from tubers supplied by the breeder, can be taken to provide rooted plantlets for pathogen testing. The breeder's stock should have been grown to a high health status and officially inspected and approved. If pathogen (mainly virus) elimination is required, larger plants can be grown from the plantlets and exposed to a heat treatment or to chemotherapy (e.g., Virozole) prior to meristem-tip culture. Dissected portions of the meristematic region of a shoot tip are placed on a liquid nutrient medium for plant regeneration. Once the resulting, or original, plantlets have been confirmed to be pathogen-free, they can be used for rapid multiplication by *in vitro* nodal cuttings under artificial, aseptic conditions in the laboratory. The healthy plant material is cut up into individual stem portions, each with one axial bud and an attached leaf (single-node cutting), which are placed on agar media in jars. The axial bud grows into stems with numerous buds from which more single-node cuttings can be taken. Once sufficient have been produced, they are allowed (with the help of growth regulators) to develop into fully rooted *in vitro* plantlets (transplants). These are then used by licensed commercial growers to produce minitubers (Pre-basic TC (Tissue Culture)) in a greenhouse or screenhouse with high health status. The transplants can be grown in soil where they are likely to produce from two to five minitubers per plant. However, many more (up to 40) uniform tubers can be achieved from frequent harvests when the transplants are grown in aeroponic, hydroponic or nutrient film cultures. In aeroponic culture the 'below ground' plant parts are suspended in air and intermittently misted with nutrient solution. In hydroponic culture the transplants are grown in static nutrient solution whereas the nutrient flows along the lower roots in nutrient film culture. The minitubers are then grown by officially approved commercial growers in the field to produce pre-basic (field grown) seed tubers for further field multiplication. Subsequent generations result in grades of basic seed and finally certified seed in the amount required by ware growers who produce potatoes for consumption. Most European countries plant whole seed tubers whereas cut tuber pieces are commonly used in African, American and Asian countries. There are usually from four to six field generations.

In Scotland seed crops are officially inspected twice during the growing season and tubers must also be inspected for diseases and disorders and meet the required standards before they can receive the official label required for marketing. The land used for seed crops must be free from wart disease and also tested and confirmed free from cyst nematodes. The interval between potato crops in the rotation must be 7 years for pre-basic seed crops and 5 years for basic seed crops. Certified seed can only be sold for ware, it cannot be replanted in Scotland. It is common practice for the Certifying Authority to hold *in vitro* pathogen-free nuclear stocks of cultivars

under multiplication, and also to fingerprint cultivars with molecular markers for unique identification (<http://www.sasa.gov.uk>).

Two other types of asexual reproduction are worthy of mention. Firstly, it is easy to take and root stem cuttings from potato plants and this can be a useful way to rapidly multiply potential cultivars for more extensive trialling in a breeding programme. Secondly, *in vitro* plantlets can be induced to produce microtubers in the axils of leaves of cuttings. These can be of value for germplasm conservation and for storage and exchange of germplasm, but perhaps are of most value in potato research.

Finally it is important to point out that strict quarantine procedures are required when potatoes are transferred from one country to another to prevent the introduction of diseases, particularly non-indigenous ones. Again advances with *in vitro* techniques are proving useful. Lang (2001) has described how CIP supplies its new cultivars to farmers in East Africa through a seed multiplication scheme that starts in Kenya with nodal cuttings being taken from virus-free sprouts. These are supplied *in vitro* from CIP headquarters in Lima, Peru to the Quarantine Station in Kenya.

## Cassava

The main difference from the potato is that the most important commercial product of cassava is a storage root, not a tuber, and hence cannot be used for reproductive purposes. In breeding, sexual reproduction is relatively easy to achieve and most programmes generate new genetic variation through crossing. The following information is taken from the detailed review by Ceballos et al. (2010), where information on past achievements and current goals can be found. High and stable productivity will remain keys traits for this staple crop; but there will also be increasing opportunities for cultivars suited to specific end uses, such as processing markets. The concentrations of cyanogenic glucosides in the roots frequently define the specific uses a given cultivar can have. Linamarin is synthesized in the leaves and transported to the roots where, on cell disruption, it is broken down by the enzyme linamarase to produce the volatile poison hydrogen cyanide (HCN). Breeding has tended to reduce the cyanogenic potential of cassava roots. We will now concentrate on breeding methods.

### ***Parents, Crossing and Clonal Generations***

Parental lines are still selected based mainly on their *per se* performance. Controlled pollinations can be used to generate full-sib progenies; alternatively half-sib families can be produced by insect pollination in polycross nurseries. With such natural pollination some self-pollination can occur when male and female flowers

open simultaneously on the same or different clonal plants. Deliberate self-pollinations are also made in some programmes to unmask useful recessive traits in the S<sub>1</sub> generation. After another generation of self-pollination, S<sub>2</sub> botanical seeds could be used to store and exchange desirable genes, rather than having to maintain germplasm *in vitro* or in the field. Male sterility can occur but is not a serious problem. It generally takes more than a year to obtain seeds of a planned cross because of the time required for plants to flower after planting, and the time required for the seed to mature. On average, between one and two seeds out of a possible three are obtained per pollination, compared with over 200 for potato. The botanical seed can be sown directly in the field, as done at the International Institute of Tropical Agriculture (IITA) in Nigeria, or first germinated in a greenhouse and then plants transplanted to the field when about 20–25 cm tall, as done at the International Centre of Tropical Agriculture (CIAT) in Colombia. As with potato, the vegetative multiplication rate of cassava is low, often around 5–10 cuttings from one plant. Thus it takes about 5–6 years from sowing botanical seed from cross-pollinations until the evaluation of clones in regional trials at several locations. Furthermore, undesirable variation is generated because cuttings from the mid-section of the stems usually produce better performing plants than those from either end. Root yields of plants obtained from botanical seed do not correlate well with those of plants from cuttings at later stages of the selection process due to differences in the root systems. As a consequence, selection on these plants is confined to high heritability traits such as plant type, branching habit and reaction to diseases. At IITA, for example, combined selection for resistance to Cassava Mosaic Disease (CMD) and bacterial blight () begins with about 100,000 seedlings and only about 3000 genotypes survive this first stage of single plant selection. In the absence of selection for disease resistance a programme may start with less than 50,000 seedlings, from which 2000–3000 are visually selected for a clonal evaluation trial with unreplicated plots of 7–8 plants from vegetative cuttings. Again selection is mainly visual, with the exception of dry matter content if determined at this stage. Commonly between 100 and 300 clones advance into a preliminary yield trial at one location, with three replicates of ten plants arranged in two rows of five with an empty row between plots to reduce interplot competition. Assessment of cooking quality traits also begins with the yield trials. Between 50 and 100 clones are selected for advanced trials at 1–3 locations, each with three replicates of 20 or 25 plants arranged in four or five rows of five plants. Finally between 10 and 30 clones enter regional trials at 6–12 locations, each with three replicates of 25 plants arranged in five rows of five plants. With four or five row plots, yield is recorded from the centre two or three rows and the outside ones used as source of planting material for the next season. The best clones are released as new cultivars after a few years of informal evaluation in semi-commercial evaluations with key farmers. They are also incorporated as soon as possible into the crossing blocks as parents for a new cycle of recurrent selection.

### ***Possible Improvements to Cassava Breeding***

The breeding scheme is thus mass phenotypic recurrent selection in which individual clones are evaluated and selected or discarded. No family data are used in the selection process. If data were recorded on all entries in the clonal evaluation trial, the general combining abilities of the parents could be estimated. The normal recurrent selection cycle in cassava requires about 8 years for completion. For important traits with a high heritability a rapid cycling scheme is possible. Ceballos et al. (2010) gives the example of carotenoid content, in particular the levels of pro-vitamin A carotenoids in order to improve the nutritional value of the crop. Root samples are taken and analyzed from seedling plants about a year after planting whilst leaving the plants growing in the field. Once the laboratory results become available, the plants with the highest contents can be used as parents for the next cycle. Thus a cycle time of 3 years can comfortably be achieved. Other ways for improving the efficiency of cassava breeding are discussed, including improved trialling and testing methods and the use of parents with good general combining abilities; but reduced cycle time should have the most dramatic immediate effect. Molecular and mutation breeding and genetic transformation are expected to make greater contributions to cassava breeding now that its genome has been sequenced ([www.phytozome.net/cassava](http://www.phytozome.net/cassava)), but will not be discussed here. The interested reader is referred to the review by Ceballos et al. (2010) for the first examples, such as marker assisted selection for resistance to Cassava Mosaic Disease, the development of SNP markers, and a mutation that produces high amylose starch which results in slowly digestible and resistant starches of value in diabetes management.

### ***Production of Planting Material***

In contrast to potatoes, there is generally no production of planting material for cassava that is independent from the commercial fields for root production. It is, however, recommended that an area in the production field is assigned as source of new planting material for the next cycle. Furthermore, when a new cultivar is identified or clean planting material (through meristem culture) is produced, specific multiplication nurseries are planted. Here the emphasis is on producing planting material rather than roots, and every effort is made to keep the material free of disease by crop rotation and the use of high altitude sites (>1000 masl). Any off-types are removed. Official inspections of the multiplication nursery are highly desirable. Details can be found in the review by Ceballos et al. (2010). It is recommended that planting material is taken from stems ranging from 8 to 18 months of age: younger stems are susceptible to pathogens and also to dehydration; older stems are too lignified with small amounts of food reserves. On average each stem yields 5–7 cuttings (stakes). There is no dormancy period so stakes can

be planted immediately after harvest; or stored on the ground in shade, in bundles in a vertical position with the apical portion of the stem pointing up. It is recommended that stakes are sprayed or submerged in a solution of insecticide and fungicide before planting. Rapid multiplication schemes have been developed involving both the use of one- and two-node microstakes and tissue culture procedures on small plants derived from pre-existing meristems or somatic embryogenesis. Again details can be found in the review by Ceballos et al. (2010).

## Sweet Potato

Sweet potato breeding is similar to that of cassava. Again the breeding scheme can be viewed as mass phenotypic recurrent selection in which individual clones are evaluated and selected or discarded. The following details are taken from the review by Lebot (2010). Most sweet potato genotypes flower naturally within the short days of the tropics, whereas breeders in temperate countries have to control day-length to induce flowering. Flowering and seed set is best when the average daily temperature is between 20 and 25 °C. Parents with one or more desirable traits are hybridized by honey bees in a polycross block to produce half-sib families. Ten vine cuttings of each parental clone are planted at 1 × 1 m spacing, with two cuttings per planting position, and staked for support. Hand pollination to produce full-sib families is also used but is time consuming and produces on average only two to three seeds per pollinated flower. Flowers are protected from pollination by insects by clipping the tip of the corolla to prevent them opening. The corolla of the pollinated flower is tied together so that insects cannot reach the stigma. For genetic studies, the female parent flowers are emasculated by hand to eliminate all possibility of pollen contamination. The exposed pistil is enclosed in a short length of drinking straw, folded over at the top to prevent entry of insects, and replaced after pollination for 2 days. Approximately 50 % of the pollinated flowers produce two seeds. As sweet potato seeds have a very hard coat they are scarified in concentrated sulphuric acid for 40 min and then rinsed under running water. This technique gives about 95 % germination success. Some 2000 seedlings are planted at 0.5 × 0.5 m spacing in the seedling nursery and are ready to harvest at 10 weeks when most of them are large enough to provide three vine-tip cuttings. Selection results in about 1000 clones being planted from these cuttings in single-hills (planting location) of three plants per genotype at 1 × 1 m (hill trial). Subsequent clonal generations comprise a preliminary unreplicated trial (100 clones) with single rows of six hills, two intermediate trials (25 and 13 clones) with three replicates of single-row plots, two advanced trials (7 and 5 clones) with four replicates of four-row plots, and finally on-farm trials (2 clones). Each generation clones are selected for a combination of different traits while the number of plants per genotype is increased to provide sufficient planting material for multiple-row replicated trials. As well as being multiplied as new cultivars, the best clones are used as parents in a new polycross block, some 8 years after the previous polycross. In most developing

countries, farmers manage their planting material stocks from vine cuttings and there is no specialized industry. In temperate countries planting is seasonal and advantage is taken of an important difference between storage and normal roots of sweet potato. Storage roots (the edible product) are produced by the thickening of adventitious roots, which results in a storage organ, but also the capacity to produce buds and sprouts on the root skin (unlike cassava). Healthy storage roots from selected plants are therefore buried in nursery beds and allowed to produce vines. When these are long enough (30–50 cm), they are cut at their base and planted directly into the new field. Each root can produce up to 15 plants. Tissue culture techniques are available in sweet potato for the healthy and rapid propagation of selected genotypes (Lebot 2010).

## Tropical Plantation Crops

Sugar cane (*Saccharum officinarum*) is a large perennial grass of 3–6 m in height, with thick, hard, juicy stems which contain high concentrations of sucrose. It is propagated by stem cuttings.

Bananas (*Musa* spp.) were dealt with in Chap. 8 on polyploidy as their breeding is dominated by ploidy manipulation. They will not be considered further in this chapter.

Oil palm (*Elaeis guineensis*) is the world's second most important oil crop after soybean. It is a perennial tree crop which is grown in plantations from seed, but where vegetative propagation is now possible. It has separate clusters of male and female flowers and forms fruits with three layers. The middle layer (mesocarp) comprises fibrous, oil-rich pulp (palm oil) and the inner black shell (endocarp) contains the small embryo with surrounding endosperm which also contains oil (palm kernel oil). In recent years genetically variable (mixed) plantations have been replaced by monocultures produced by the tissue culture clonal propagation technique. This is a trend that may occur in other tree crops. However, oil palm will not be considered further here as breeding for clonal propagation has not been the main method; but the interested reader is referred to the review by Soh et al. (2009).

## Sugarcane

Sugarcane is native to southeastern Asia, with its cultivation in India dating from 5000 years ago. It has C<sub>4</sub> photosynthesis, resulting in a vigorous biomass accumulation under tropical conditions, but less vigorous growth in temperate regions. In commercial settings, sugarcane is clonally propagated via stem cuttings. The plant is semi-perennial, with a period of 12–18 months from planting to first harvest under tropical conditions. Ratoons (sugarcane stumps) regrow after each harvest, but with decreased vigour, so that on average five productive harvests can be

achieved from a plantation before replanting is necessary. Sugarcane stores energy as the non-reducing disaccharide sucrose, which accumulates in large amounts in the vacuoles of parenchyma cells of stem tissues (up to 23 % w/v). Sugarcane is responsible for about 70 % of the world's raw table sugar (sucrose) production with the rest coming from sugar beet in temperate countries. In 2011 (<http://faostat.fao.org/>), 1800 million metric tonnes (MT) of sugarcane (not sugar) were produced worldwide from a total area of 25.4 million hectares (FAO), with the largest producer Brazil (734 million MT) contributing 41 % of the total, followed by India (342 million MT) and China (115 million MT). Production in Brazil is of particular interest because of its importance as a biofuel for energy production as well as for raw sugar. The following account is taken primarily from the review by Scortecci et al. (2012). More information can be found in an earlier and more extensive review by Ming et al. (2006).

According to the Brazilian Ministry of Mines and Energy, energy production in Brazil in 2008 was 46 % renewable, with 15 % derived from hydroelectric, 16 % from sugarcane, 12 % from coal and wood and 3 % from other sources. In sugarcane refineries, 20 % of the energy used in boilers came from *bagasse*, the remaining fibrous matter after juice extraction from sugarcane stalks. Ethanol derived from sugarcane is currently responsible for about 50 % of vehicle fuel in Brazil, where a biofuel programme has been in place since 1975 with the successful aim of relieving the nation's external dependence on oil. Sugarcane is considered a first-generation biofuel crop. In Brazil, most energy converted to ethanol biofuel is derived from the sucrose extracted from squeezing the stems and fermenting the juice. Hence dual purpose cultivars could be bred with high sucrose content and yield for both sugar and alcohol production. However, only a third of the plant's energy is extracted using this technology, with the remaining being stored in less readily available compounds such as the cellulose deposited in plant cell walls along with hemicellulose and lignin. The second generation of biofuels requires the development of an efficient process of breaking down cellulose (thermally or enzymatically) into fermentable sugars from the *bagasse* and leaves. It also requires cultivars bred specifically for this purpose, ones which use metabolic energy to accumulate either sucrose or cellulose for maximum biomass production per unit of area. The interested reader is referred to Scortecci et al. (2012) for a discussion of breeding objectives in sugarcane for both first and second generation biofuels, as well as for raw sugar. In addition to defining objectives for target end uses and target environments, two objectives relate to symbiotic associations and hence to farming systems in terms of inputs. Sugarcane is able to establish a mutualistic symbiosis with diazotrophic endophytes (bacteria) that reside in xylem cells and in the rhizosphere, and which partially supply the plant's nitrogen requirement. Sugarcane is also able to establish an association between arbuscular mycorrhizae (AM) species and its roots, which improves the plant's nutrient uptake, notably phosphate, but also nitrogen and possibly water and other nutrients.

## Parents, Crossing and Clonal Generations

The history of sugarcane breeding, from intraspecific crossing (open-pollination) of noble canes (*S. officinarum*) at the end of the nineteenth century to modern cultivars, was dealt with in Chap. 8 and will not be repeated here. Modern cultivars are complex hybrids synthesized principally from noble canes and *S. spontaneum*, with minor contributions from *S. sinense* and *S. barbieri*, and to a lesser extent, *S. robustum*. New cultivars are derived from extensive recombination and selection among materials of this kind.

Hybridization is used to create genetic variation from which superior genotypes are selected over clonal generations; the whole process taking 12–13 years, as in potatoes. As explained in Chap. 8, modern cultivars are complex hybrids from an inter-breeding group known as the ‘*Saccharum Complex*’. Biparental crosses among them are preferred to polycrosses as the breeder has control over the choice of both parents, which can be chosen for trait complementarity. When sugarcane flowers, the plant stops growing, and sugars accumulated in the stalks are used for reproductive development, followed by plant senescence. Flowering is regulated primarily, but not exclusively, by photoperiod with short days of 11.5–12.5 h required for induction. In Brazil, parental plants are therefore cultivated and crossed in tropical breeding stations that meet these requirements. Subsequent progeny (clonal) evaluation is done in the major subtropical production regions that do not favour flowering. In temperate climates, hybridizations can be done in a crossing house in which temperature and humidity are controlled as well as photoperiod.

Since sugarcane flowers are hermaphrodites, emasculation of the female parent is required to avoid pollen contamination; for example, with heat treatment of the panicle (arrow) by immersion in water at 50 °C for 4.5 min. However, emasculation is not always done due to the large number and small size of flowers on the arrow and to fact that it can take 14 days for the entire arrow to complete flowering. Some self-pollinated seed can then result in the absence of self-sterility. Should self-pollination actually be required in a breeding programme, the arrow would need to be covered, usually with a paper bag. Flowering stalks of the chosen male parent are cut and labelled, transported to the crossing shed, and placed slightly above the female, as pollination occurs by gravity. The two parents are protected in a ‘lantern’ to avoid cross pollination, i.e., a transparent cylinder with closed top and open bottom hung over them. Alternatively, isolated cubicles in the crossing house can be used. The stalks are kept in a nutritive solution, which is replaced frequently to preserve the stalks for about 25 days. The panicles are dried in a controlled room at 32 °C and low humidity for 3 days, after which the seeds are ready for sowing. If necessary, seed can be dried and stored at temperatures of 0–5 °C.

In a typical breeding programme described in Sleper and Poehlman (2006), 100 crosses are made each year and 1000 seedlings raised from each cross, making a total of 100,000, which is comparable to a potato breeding programme. Seed is germinated in a greenhouse and seedlings transplanted to the field. Individual plants are selected for traits with a high heritability. For most of the traits of commercial

importance, individual plant selection is ineffective due to high environmental variation. However, family selection for these traits can be effective and Cox and Hogarth (1993) concluded that family selection with replication of each family, followed by individual plant selection within the best families is the most efficient way to select sugarcane. Selected seedling plants (10,000) are used to initiate multitrait, multistage clonal selection in field experiments which can take 10 years until an elite genotype is released as a new cultivar. Breeding clones can be propagated by single-bud cuttings with several made from each selected clone. Thousands can be ‘germinated’ in relatively small areas of greenhouse benches. Selected clones move, with multiplication, from unreplicated canes (10,000) to unreplicated plots (1000), through replicated local trials with one ratoon crop harvest to replicated regional trials with two ratoon crops harvested, and then final characterization and release of superior clone. As selection advances, the number of genotypes decreases, allowing an increase in replication number, plot size and experimental locations. At later stages, the clones are also evaluated for diverse harvesting time, if the local cropping management system includes more than one harvest a year. Further details, particularly variations on practical ones, can be found in the chapter on Breeding Sugarcane in *Breeding Field Crops Fifth Edition* by Sleper and Poehlman (2006).

## Olive Breeding

Olive (*Olea europaea*,  $2n = 2x = 46$ ) breeding is included briefly as an example of an oil fruit tree, second in importance after oil palm, and traditionally cultivated in the Mediterranean area since 3000 BCE. Details of the olive’s origin, cultivation and traditional breeding can be found in the chapter by Baldoni and Belaj (2009) in Volume 4 of the *Handbook of Plant Breeding*. The first cultivars were probably selected from trees bearing large fruits and/or high oil content and were vegetatively propagated, either via cutting or grafting onto indigenous oleasters (wild olives). Fruit production starts 3–5 years after planting and olive orchards can survive almost indefinitely due to the longevity of the trees, which is typically over 500 years, but in practice to no more than 50–60 years. As a consequence, there may have been relatively little selection since domestication and probably only a few generations separate the presently cultivated forms from their progenitors. Until recently there have been few efforts to produce new cultivars by sexual hybridization, partly because under normal conditions olive seedlings don’t begin to set fruits until 15–20 years after germination. In contrast, new methods of cultivation have been developed based on the complete mechanization of harvesting and pruning. Flowers are generally hermaphroditic and wind pollinated, and most cultivars are self-incompatible. The olive fruit is a drupe and the oil accumulates in its mesocarp rather than the seed. Virgin olive oil is mechanically extracted. The major oil (triglyceride) is mono-unsaturated oleic acid, but a vast array of microconstituents are also present and thought to contribute to its beneficial

effects on human health. As a consequence, DNA fingerprinting is used to check the provenance of virgin olive oil. Classical breeding by hybridization and selection is relatively new, and not surprisingly there has been an emphasis on shortening the juvenile period. Inter-cultivar crosses can be made by adding pollen of the paternal parent to bagged branches of the maternal one chosen from among self-sterile cultivars.

The initial selection steps are still being researched, as seen in the paper by León et al. (2015) in which they demonstrate the utility of an intermediate step of selection between the seedling trials and the final multi-locational trials. Seedlings (1545) were secured from crosses and open pollinations of 12 different parents made in 1998 and 1999. They were submitted to a forcing growth protocol in a greenhouse for 18 months, transplanted into a field at 1.5 m × 3 m spacing, trained to form a canopy at 160 cm height, and then allowed to develop freely. Plants reached their adult phase and started flowering during their second year in the field, and were assessed for two further years. Semi-hardwood stem cuttings were taken in 2006 from the 109 most promising plants (genotypes) and the propagated trees were planted in the field in 2007 in a randomized block design with four replicates of single tree plots. Selection had been mainly on the basis of early flowering and crop (i.e., short juvenile period), high oil content (fourth year harvest) and growth habit (trunk diameter in fourth year). The same assessments as before were made during the second, third and fourth years after planting in the field. In addition, oil quality components were evaluated at two harvest dates (mid September and mid December) during the fourth year. As a result of this second step of selection, 14 genotypes were selected for final evaluation on the basis of their overall performance. The authors concluded that selection for fruit size and oil content, but not plant vigour and visual assessment of crop yield, had been effective in the seedling population. Furthermore, the high number (109) of genotypes retained in the intermediate stage meant that variation remained and selection could be practised for fatty acid composition, tocopherol and phytosterol contents, and phytosterol profile; traits with high to moderate heritability at this stage. The authors concluded that this new selection scheme was superior to the one they used in their breeding programme started in Spain in 1990 in which one new cultivar was selected from 15 promising seedlings out of an initial population of 748 (i.e., 2 % selected). The 15 genotypes were assessed in a randomized block design with 16 single-tree replicates of each.

## Fruit Breeding

It is not possible to do fruit breeding justice in this general book on plant breeding. The interested reader is referred to the 875 page Volume 8 in the Handbook of Plant Breeding (Badenes and Byrne 2012). I have chosen strawberries, raspberries, grapes and apples as my examples, the latter in preference to peaches (*Prunus persica*) which are often considered a model fruit species for genetics and

genomics. This is because the peach has a shorter generation time and smaller plant size than other major fruit crops, is a diploid species with a small number of chromosomes ( $2n = 2x = 16$ ), is self-fertile and tolerant of inbreeding. A high-quality draft genome of peach was published in *Nature Genetics* online on 24 March 2013 (Verde et al. 2013). However, worldwide apples are economically more important and allow me to illustrate typical issues in tree fruit breeding.

## **Strawberry Breeding**

The following brief account of strawberry breeding is based on the reviews by Hancock et al. (2008b) and Chandler et al. (2012). The diploid wood strawberry species *Fragaria vesca* was probably cultivated by the ancient Greeks and Romans, and by the 1300s was being grown across Europe. Today's principal cultivated 'pine' strawberry (*Fragaria × ananassa*), however, originated in France during the eighteenth century when female clones of dioecious *F. chiloensis* from Chile were interplanted with pollinator clones of *F. virginiana* that had previously been imported from eastern North America (for full story see Kingsbury 2009). Hybrid seedlings appeared that combined the relatively large fruit size of the former species with the hardiness and productivity of the latter species. Although both progenitors are octoploid species ( $2n = 8x = 56$ ), recent research suggests that their hybrid is largely diploidized through its subgenomes and that inheritance is mainly but not completely disomic (Rousseau-Gueutin et al. 2008). Furthermore, nearly all modern cultivars are hermaphroditic, the first of which was 'Wilson's Albany' which James Wilson of Albany, New York produced by selection and released in 1852 (Kingsbury 2009). They have the great advantage of producing fruit on self-pollination. Today the strawberry is the most widely distributed fruit crop in the world, being grown in most countries with a temperate or subtropical climate, and also in the highland areas of many tropical countries. The world's largest strawberry-producing country is the USA where over 80 % of the crop is consumed fresh.

Strawberry plants are easily propagated vegetatively as they produce runners (stolons) on which daughter plants develop. Although the strawberry is a short-statured perennial plant, it is increasingly being cropped for less than a year, and at most for 3–4 years, because young plants tend to have larger, higher quality fruit. The crop can be grown in open fields or in polyethylene tunnels for protection, either on the ground or in raised hydroponic channels. Plants may be spaced and runners removed so that crowns are the primary yield component, or grown in matted rows in which runners provide most of the crop. There are two main types of strawberries, ones which initiate flowers in short days (<14 h) and ones that are day-neutral provided the temperature is below a critical maximum ( $\leq 25^{\circ}\text{C}$ ).

Breeding activities started in Europe in the early 1800s, particularly in Great Britain from 1817 where Thomas Andrew Knight used clones of both *F. virginiana* and *F. chiloensis* in his crosses. In North America, Charles Hovey produced the first

important American cultivar, 'Hovey', by crossing the European pine strawberry, 'Mulberry', with a native clone of *F. virginiana* in 1836. Extensive breeding followed on from these pioneers. By the middle of the twentieth century much breeding work was being done in Europe and North America in the public sector. Today there are numerous public and private breeding programmes across the world. A typical breeding programme has been described by Chandler et al. (2012) in which 100 different controlled crosses are made, from 30 or more heterozygous parents to combine their desirable phenotypic characteristics in the resulting offspring. The female parents are grown as potted plants in a greenhouse. Crossing is done by brushing the anthers of a detached flower of the male parent onto the receptacle of an emasculated flower of the female parent with its numerous pistils. Seeds from fruits of the same cross are extracted in water and dried. They are scarified (for water penetration) before sowing in seed trays. After about 6 weeks the resulting seedlings are transplanted into peat pellets and grown for another 6–8 weeks before being planted in a field nursery. One hundred seedlings are raised from each cross, making a total of 10,000. Fruits from these seedlings are evaluated in their first year in the nursery and again in their second year. Several hundred seedlings (genotypes) are selected in summer and runners collected in autumn to provide ten or more plants of each genotype for subsequent evaluation in (one or two) small plots. A dozen or more of these are selected and multiplied for assessment in replicated trials. Initially these are done at one site, but subsequently over sites and seasons before deciding which potential cultivars to commercialize. These are also used as parents in the next cycle of crossing and selection.

## Raspberry Breeding

The European red raspberry (*Rubus idaeus*) was first mentioned by the Roman author Pliny the Elder (Kempler et al. 2012) and was cultivated all over Europe by the 1500s. Modern cultivars are derived from natural hybrids of the European and North American subspecies *R. idaeus* var. *vulgatus* and *R. idaeus* var. *strigosus* dating from the early nineteenth century, after the introduction of the latter to Europe. Initially new cultivars were selected from self-sown seedlings, but deliberate hybridizations were made by breeders from early in the twentieth century. The history of improvement can be found in the review by Finn and Hancock (2008).

The red raspberry is a deciduous shrub with perennial roots and biennial stems (floricanes) which produce fruit in their second year, after a dormant period prior to flowering. However, modern cultivars have been bred that are autumn fruiting because they produce fruits on their primocanes in the first year of growth. The red raspberry is widely grown in all temperate regions of Europe, Asia and North America. It is a diploid ( $2n = 2x = 14$ ) protandrous outbreeding species in which self-pollination can occur. Hence they can produce fruit on self-pollination. Emasculation of the female parent is required in controlled hybridizations, usually done in a glasshouse, and flowers of both male and female parent are commonly bagged

to prevent any contamination from unwanted pollen. Seeds are extracted from the ripe fruits and stored at room temperature for a few months before sowing. They require both scarification (for water penetration) and stratification (to simulate low temperature, winter conditions) for germination.

A typical breeding programme has been described by Kempler et al. (2012). Annually around 100 seedlings are raised from each of 25–60 crosses and transplanted into pots in a greenhouse. Here they can be screened for the spineless trait and resistance to pests and diseases such as aphids and root rot. On average 30 % of seedlings are discarded before planting. Hence around 70 seedlings per cross are planted outdoors in spring at wide spacing: 75–150 cm within rows 240 cm apart. A total of between 2000 and 5000 field-grown plants is therefore typical. Intense selection (1 %) takes place in the second year after planting. Leaf tissue from each selection is tested for the presence of virus and the plant discarded if the test is positive. Uninfected plants are vegetatively propagated for further evaluation, and may also be transplanted to another field for use as parents and further testing for virus. Propagation can be done by simple division of a plant into single canes in autumn or by root cuttings. If roots are cut into short sections in autumn (say 2 cm long), the shoots which arise from them can be propagated (rooted) as stem cuttings under mist. A further possibility is propagation over winter by tissue culture from stem nodes from the primocanes. First year trials include standard cultivars as controls and are typically done in one of two ways. In breeding for the fresh market, the trials have three replicates of three-plant plots and 2 years after planting are hand harvested for fruit evaluation. In contrast, in breeding for the processing market, ten plants are grown in unreplicated plots and machine harvested for assessment in the second and third year after planting. Promising clones are propagated and planted in large-scale growers' trials for further evaluation and also used as parents in the next cycle of breeding.

## **Grape Breeding**

The following brief account of grape breeding is based on the reviews by Owens (2008) and Reisch et al. (2012). The grapevine (*Vitis vinifera*) is a woody, deciduous, perennial climber with coiled tendrils, large, toothed leaves and clusters of inconspicuous flowers that develop into bunches of juicy berries. It is one of the oldest and most valuable horticultural crops in the world, now grown on about 7.9 million hectares. Grapes are processed into wine and non-alcoholic juice, dried into raisins, and used for fresh consumption (table grapes). Viticulture and wine production were an important part of the cultures of ancient Egypt, Greece and Rome. Today grapevines are grown in Europe (particularly France, Italy and Spain), North and South America (USA, Argentina and Chile), Australia (Fig. 11.7), South Africa and China; in regions with hot, dry summers and cold, wet winters. Wine grape cultivars usually have relatively small and seeded berries whereas table grape cultivars have relatively large seedless berries. Centres of diversity of wild species



**Fig. 11.7** Balnaves Winery, Coonawarra, Australia (18 September 2013)

of *Vitis* occur in North America (30 species) and East Asia (30 species), but the most widely cultivated species, *Vitis vinifera* ( $2n = 2x = 38$ ), is the sole native species in Europe, the Near East and Northern Africa. It was domesticated from *Vitis vinifera* subsp. *sylvestris*, first some 6000–10,000 years ago in the mountainous regions between the Caspian and Black Seas. From there cultivation spread to the east and west, including southwards to the western side of the Fertile Crescent, the Jordan Valley, Egypt and eventually throughout the Mediterranean. However, the natural range of *Vitis vinifera* subsp. *sylvestris* was extensive, and there may have been two or more secondary domestication events. The species can certainly be divided into three morphotypes: Western European *occidentalis* with small clusters of small berries, Central Asian *orientalis* with loose clusters of large berries, and intermediate Eastern European *pontica*. Furthermore, molecular genetic analysis of the US Department of Agriculture (USDA) germplasm collection of *V. vinifera* and *V. vinifera* subsp. *sylvestris* has provided evidence for introgression of local Western European *sylvestris* into Western European *vinifera* cultivars (Myles et al. 2011). Domestication resulted in greater uniformity of berry maturity within clusters, higher sugar content and a wide range of fruit colours.

*Reproductive biology:* As with strawberries, we have another example of dioecious wild species but a domesticated crop with perfect flowers so that fruit can form following self-pollination. The genetics of sex expression appears simple, a single locus with staminate flowers dominant to perfect flowers which in turn are both dominant to pistillate flowers. Reproduction by self-pollination would result in

severe inbreeding depression, but the crop is normally propagated by vegetative reproduction; commonly today by hardwood cuttings either outdoors or in a greenhouse. All grapevine cultivars are in fact highly heterozygous with as much as 13 % sequence divergence between alleles. Hence genotype PN40024 (derived from Pinot Noir) was chosen for sequencing a grapevine because it had been bred close to full homozygosity (about 93 %) by successive selfings (Nature, 26 August 2007). This permitted high-quality whole-genome shotgun assembly. Two key features of cultivation and domestication of grapevine by vegetative propagation have been stable bud sports and virus infection. Bud sports are somatic mutations that can be propagated and recognized as clones and cultivars, and can sometimes be chimeras. Virus infection can be a non-genetic source of clonal variation. Today cultivars may in fact be a mixture of clones.

*Modern breeding:* Modern breeding using deliberate hybridizations started in Southern France in 1824 and in North America in 1830, the latter with crosses between European and American vines. The further development of modern breeding was intimately associated with the arrival of North American diseases and pests in European vineyards in the mid nineteenth century. The root louse, phylloxera (*Daktulosphaira vitifoliae*), powdery mildew (*Uncinula necator*), downy mildew (*Plasmopara viticola*) and black rot (*Guignardia bidwellii*) all caused substantial losses in the highly susceptible *V. vinifera* vines of Europe. Two important advances in breeding occurred as a result of these epidemics. The first was the advent of rootstock breeding as an effective and immediate means to control phylloxera. Use of the horticultural skill of grafting allowed important susceptible European cultivars to be grown as scions on rootstocks of resistant wild vines imported from North America, initially dormant cuttings *V. riparia* and *V. rupestris* as they root easily. Subsequently hybrids of these species with *V. berlandieri* were used as the latter provides adaptation to calcareous soils (high pH). Resistance to nematodes has been identified in many wild species and incorporated into rootstock breeding programmes. Rootstocks that are tolerant of virus infected scions have also been bred. The second important advance was breeding programmes to combine resistance to phylloxera with resistance to fungal pathogens in a single cultivar. These were begun as early as 1874 and resulted in hybrid direct producers (HPDs), the first of which were imported to Europe from the USA. The cultivars were primarily hybrids of wild *V. labrusca*, *V. aestivalis* and *V. riparia* with *V. vinifera* cultivars. However, breeders in France used *V. rupestris* rather than *V. labrusca* because the latter was associated with unpopular flavours. Even so, the French still considered the flavour of these hybrids inferior to the ‘true grape’, and hybrids were banned from quality wines from 1927 (Kingsbury 2009). Hence hybrids became popular in France for table rather than quality wine, but eventually this use was discouraged from 1953 by further laws restricting their planting and sale. As a result, traditional *V. vinifera* grapes grafted onto hybrid rootstocks came to dominate. Despite much activity worldwide in breeding interspecific hybrid scion cultivars, today, *V. vinifera* grapes still constitute more than 95 % of the world market, and there has been notable success in the development of new

cultivars. There is, however, much interest in backcrossing disease resistance genes into *V. vinifera* from other species, while separating them from low fruit quality genes typical of these other species. Because grapes are highly heterozygous, different recurrent parents are used to avoid inbreeding depression.

*Breeding methods and techniques:* Breeding methods and techniques can be summarized in brief as follows. Pair crosses are made between (elite) parents with complementary desirable phenotypes (and genotypes) and large seedling populations raised. Controlled pollinations are done by cutting a previously bagged, freshly blooming cluster and tapping it lightly against an emasculated cluster, which is immediately bagged again. Pollen that has been stored can be applied with a camel's hair brush. Emasculation is necessary since hermaphrodite grapes are self-fertile. Viable grape seeds can be extracted from mature fruit in water. Cold stratification at 0–10 °C under moist conditions for about 3 months is required for quick, uniform and high germination. Seedlings are raised in a glasshouse with supplementary lighting (16 h) if necessary and daytime temperatures of 28–32 °C followed by 22 °C at night. Seedlings are either planted outdoors in permanent vineyard locations (own-rooted) or first grown in a field nursery for 1–2 years to attain sufficient size for successful transfer to the vineyard. Grapes have a juvenile period of 2–4 years after which evaluations can be made for a number of years. Once plants with potential have been identified, they can be propagated to other vineyards and other locations for replicated testing, including grafting to desirable rootstocks. Breeders at many locations in North America and Europe typically test wine potential using microvinification techniques that start with the grapes available from the original seedling vine and later from multiple vine plots.

During the last 50 years the table grape market has grown rapidly with Japan and Korea leading the way in the development of large-berried seeded and seedless cultivars. The former are usually autotetraploid cultivars from  $4x \times 4x$  crosses and the latter triploid cultivars from  $4x \times 2x$  crosses. Most of these Asian grapes are derived from both *V. vinifera* and *V. labrusca*. Interestingly, in eastern North America, as well as Asia, the fruity and aromatic flavours of *V. labrusca* are appreciated. Embryo rescue and culture enables crossing between seedless grapes and results in a very high percentage of seedless progeny. Seedlessness is also important in raisin grape breeding.

## **Apple Breeding**

In considering the grape we came across grafting onto a rootstock as a way of controlling soil pests and hence the need to consider scion and rootstock breeding. Grafting is common practice with tree fruits where the rootstock can have a major influence on the control of tree size, as happens in our chosen example of the apple. The following information is taken primarily from the reviews by Hancock et al. (2008a) and Brown (2012).

Apples are grown in most temperate climates and require a period of cold ( $<7^{\circ}\text{C}$ ) to flower and grow normally. Hence projected climate change is likely to pose new challenges for apple producers. Currently China is the largest producer of apples, followed by the USA and countries in the European Union, particularly Poland, Italy and France. *Malus sieversii* from the mountains of NW China is the probable progenitor of domesticated apple through hybridization with wild species to the east and west. The result was an interspecific hybrid complex, designated *Malus × domestica*, but sometimes the domesticated apple is referred to as *M. pumila*. Selected cultivars were established from these hybridizations and disseminated through grafting. There are reports of apples from 4000 BCE and later Roman authors documented apple culture. Indeed, the Romans spread the apple across Europe during their invasions.

The cultivated apple can be regarded as an allopolyploid that displays disomic inheritance ( $2n = 34$ ). Self-pollination is largely prevented by a gametophytic self-incompatibility system so that pollinator trees are required in modern orchards comprising clonal cultivars; pollination is done mainly by bees. Triploids can occur spontaneously in crosses between diploids and some popular cultivars are triploids with large fruit, but cannot supply pollen for fertilization because they are male-sterile. Settlers in North America brought apple seeds and grafts with them that were not well adapted to the “new world”; but they soon learned to propagate the seedlings best adapted to the new climate. Consequently thousands of new cultivars were established and named. However, again it was Thomas Andrew Knight in Great Britain at the beginning of the nineteenth century who is credited with being the first apple breeder to use deliberate hybridizations. The late nineteenth and early twentieth century represented the maximum of diversity in apple cultivation in Europe with hundreds of locally popular cultivars being grown in thousands of small orchards. Today, at the beginning of the twenty-first century, just a few cultivars dominate world production and are widely used in breeding programmes around the world.

Rather than simply describe a modern breeding programme, we are going to look at the potential of genomic selection. The reader interested in breeding objectives and achievements can find them in the review by Brown (2012). The reader can also get a feel for a real breeding programme from the paper by Durel et al. (1998) on the utilization of pedigree information to estimate genetic parameters, in a French programme aimed at combining disease resistance and fruit quality. Over the period 1974–1988 the breeders made 210 controlled crosses and raised progenies (full-sib families) of sizes from 2 to 238 (mean 61.9). One flower may only produce one seed due to damage from emasculation, so 50–100 pollinations are typically required to generate a few hundred seeds. Seedlings (12,900) were assessed for scab (*Venturia inaequalis*) resistance in a glasshouse test and approximately 40 % selected and progressed to a nursery for screening for resistance to powdery mildew (*Podosphaera leucotricha*). After two growing seasons approximately 30 % were selected and progressed to experimental field plots for evaluation of fruit quality and other traits. The experimental field was planted over the period 1978–1990.

## ***Genomic Selection in Apple Breeding***

The use of genomic selection for complex (polygenic) traits is theoretically attractive in tree fruit breeding. The apple programme at Plant & Food Research (PFR) in New Zealand provides a good example of what is currently being done experimentally (Kumar et al. 2012). Genomic selection became possible with the discovery of more than two million single nucleotide polymorphisms from sequencing the complex heterozygous genome of the apple cultivar ‘Golden Delicious’ (Velasco et al. 2010), and from whole genome low-coverage re-sequencing of other germplasm including some of the commonly used founders in worldwide apple-breeding. This enabled the development of an 8000 SNPs Illumina Infinium assay for use in genotyping.

The PFR programme consists of parental breeding and cultivar development. The former comprises cycles of crossing and selection in a population improvement scheme which includes input of new germplasm. Each cycle provides parents for the cultivar development programme. These are crossed to produce large seedling families for assessment and selection (stage 1). Then the most promising seedlings are cloned by grafting onto a clonal dwarfing rootstock (e.g., ‘M.9’) for trialling across different environments (stage 2), followed by larger-scale testing of the best selections in commercial orchards (stage 3). It generally takes about 6 years from seed before outstanding individuals can be identified to progress to stage 2, or be used as a parent in another cycle of crossing. This is because apples have a long juvenile period where even on dwarf rootstock, significant fruiting does not generally occur until the fifth year after seed germination. Implementation of genomic selection to identify and select outstanding seedlings much earlier (first year of growth) would therefore increase the efficiency of the breeding system. Genomic selection utilizes all available genome-wide markers simultaneously to predict breeding values (potential parents) and total genetic values (potential cultivars).

PFR generated a training population in 2004 to implement GS in its breeding programme. The population of 1400 individuals from eight full-sib families was evaluated for fruit quality in 2010, and genotyped along with the parents and grandparents using the Illumina Infinium 8 K SNP chip. SNP effects are being estimated in 1200 randomly selected seedlings and cross-validated in the remaining 200 seedlings. Genomic selection will then be practised on a selection population of 2000 young seedlings from 10 crosses with pollen parents from the training population and seed parents identified from previous progeny trials. In other words, the two populations will be genetically related. The effectiveness of the selection will be evaluated in 2016 when phenotypic data becomes available. The interested reader can find more details of the experiment in the review by Kumar et al. (2012).

## Clonal Cultivars Produced by Apomixis

We will finish this chapter with an example of clonal cultivars produced by apomixis in bluegrasses (*Poa* species), taken from the review by Huff (2010). In temperate climates, bluegrasses are utilized as components of pastures, meadows and cultivated turfs. Most species possess hermaphroditic flowers but gynomonoecious (female and hermaphrodite flowers occur separately on same plant), gynodioecious (female and hermaphrodite flowers occur on different plants) and dioecious species also occur. Modes of reproduction include self-pollination, cross-pollination and facultative apomixis. Kentucky bluegrass (*Poa pratensis*) is the botanical-type species for the genus *Poa* and is also the most economically important and widely utilized bluegrass species for both forage and turf. The breeding of Kentucky bluegrass is governed by its form of apomixis, known as facultative pseudogamous apospory, which is simply inherited and environmentally stable. A nucellar cell mitotically develops into an unreduced embryo sac (apospory), the nucellus being the mass of tissue in the ovule that contains the embryo sac. The unreduced embryo sac continues development autonomously, without pollen fertilization of the egg (parthenogenesis by pseudogamy), but requires pollen fertilization of the central cells for proper endosperm development. In the absence of meiosis, apomictic plants are capable of harbouring and propagating odd ploidy levels and aneuploid chromosomal abnormalities, and chromosome numbers in Kentucky bluegrass range from 18 to 156. The apomixis is facultative as various components of the process occasionally break down, resulting in progeny plants that are distinctly different from the seed-bearing parents; that is off-types or aberrants. In natural grassland systems, successful genotypes of Kentucky bluegrass often dominate an ecosystem aided by the true-to-type apomictic breeding system. Before 1970, all cultivars of Kentucky bluegrass were initially collected from pre-existing naturalized grasslands or cultivated turfs. Cultivar 'Kenblu' was released in the 1950s and is still in use today for pastures and as a low maintenance turf. Its success may partly be due to the value of maintaining genetic diversity within a grass sward. It was originally developed as a blend of different naturally occurring apomictic ecotypes that shared similar growth performances and phenologies, and which can now be grouped into three main clades using DNA markers.

Since 1970, most cultivars have been derived from intraspecific hybridizations for which two strategies exist. The most common one is to make pairwise crosses between parental genotypes that are each highly apomictic and then look for rare aberrants. Large numbers of progenies have to be evaluated visually, or screened using molecular markers or flow cytometry. However, the rare aberrants are likely to exhibit the high levels of apomixis (95 % or more) required to give a new cultivar uniformity and stability across generations of seed production. In addition, crossing techniques have been discovered that significantly increase the number of aberrant offspring derived from a highly apomictic female parent. One technique is to bring field grown and vernalized plants into a greenhouse to flower under artificial long daylengths, and then to pollinate as soon as possible after the stigmas emerge and

are receptive. Mechanical means can be used to achieve a large number of pollinations. Inflorescences are bagged in a ratio of five male to one female to ensure pairwise crossing. The pollination bags are then attached to a wire rack with a wooden frame suspended from ceiling chains so that it can be regularly nudged by an electric motor. The other strategy for making intraspecific hybridizations is to use parents (typically the female) that are highly sexual in order to increase the frequency of aberrant recombinant progeny from sexual reproduction. The downside of this strategy is that levels of apomixis among the progeny tend to be low; so once again it is necessary to evaluate large numbers of progeny, this time to identify those with high levels. Most often, these high levels appear in the generations succeeding the  $F_1$  as genetic elements controlling apomixis are recombined.

## Conclusions

It is probably fair to conclude that the multistage, multitrait selection of clonal cultivars from large seedling populations (10,000–100,000 in size) in many crops has not been guided as well as it might have been by theoretical considerations. Furthermore, one is left wondering the extent to which selection of seedlings has been more successful than Finney's suggestion of discarding at random before doing any assessments, so that those actually done allow effective selection. However, more encouragingly for the future, there is the potential for the breeding of clonally propagated crops to benefit enormously from genomic selection of seedlings in order to save time and resources.

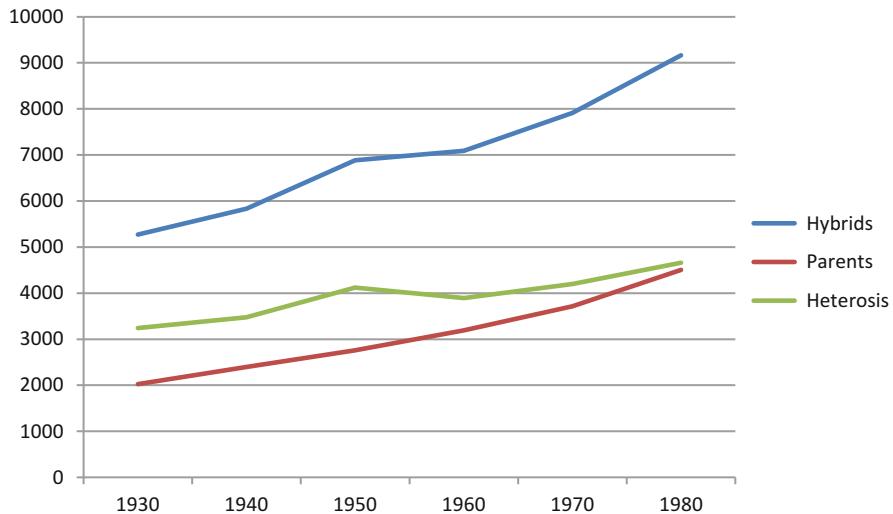
# **Chapter 12**

## **Hybrid Cultivars from Inbreeding and Crossbreeding**

### **Introduction**

In the last chapter we saw how a desirable genotype could be found by multi-trait, multistage selection over clonal generations in crops that can reproduce by vegetative reproduction. In outbreeding crops that reproduce by sexual reproduction, the selection and multiplication of desirable genotypes is done indirectly through inbreeding and crossbreeding. Multi-trait, multistage selection takes place during the inbreeding process, but selection of the inbreds is ultimately on their ability to produce the desired genotype in a hybrid. Hybrid maize was introduced in Chap. 2 as one of the great success stories of plant breeding in the twentieth century. Now it is time to consider the theory behind hybrid breeding and its application to maize and other crops. First, however, it is worth recalling some essential features of the hybrid maize story in the USA.

In 1904 East noticed that all of the high protein ears in the Illinois Long Term Selection Experiment (ILTSE) traced back to a single ear of Burr's White and that the other strains also had narrow pedigrees, and hence they were inbred. He reasoned that deliberate inbreeding and selection should be a faster way to homozygous lines with desirable chemical compositions. East left Illinois for Connecticut Agricultural Station in 1905, but subsequently he and his students selected inbred lines from a wide variety of germplasm, including Burr's White from Illinois. East (1908) and Shull (1908, 1909) showed that homozygous and homogeneous pure lines (inbred lines) could be developed from an open-pollinated maize cultivar by five to seven generations of selfing. The inbred lines were often weak and difficult to propagate, a phenomenon known as inbreeding depression. When certain of the weak inbred lines were crossed, vigour was restored (hybrid vigour) and the yield of the inbred line crosses (single crosses or  $F_1$  hybrids) usually exceeded that of the original open-pollinated cultivar from which the lines were developed, a phenomenon known as heterosis. The difficulty of propagating weak inbred lines appeared to prevent commercial exploitation; but Jones (1918) solved the problem by crossing two vigorous single crosses to



**Fig. 12.1** Increases in maize yields ( $10,000 = 10 \text{ t/ha}$ ) of inbred parents and their single-cross hybrids available from 1930 to 1980, where heterosis is difference in their yields (data from paper by Troyer (2006))

produce double-cross (DC) hybrid seed for farmers to grow, and showed that these were also superior (often by 20 %) to their open-pollinated cultivars. Initially inbred lines were developed from the best open-pollinated cultivars, but after 1940 further progress was made by crossing the elite, inbred lines which were now available (Hallauer and Carena 2009). Lines developed from different cycles of recurrent selection in BSSS (Iowa Stiff Stalk Synthetic, Chap. 10) have also been a major contributor of germplasm to newer inbred lines (Hallauer and Carena 2009). Eventually the inbred lines were good enough for the commercial production of single-cross (SC) hybrids. Recent experiments with historical (SC) hybrids and their inbred parents provide evidence of a 122 % increase (from 2.03 to 4.51 t/ha) in the yields of the parental inbreds available from 1930 to 1980 and a corresponding 74 % increase (5.27 to 9.16 t/ha) in the yields of their hybrids (Troyer 2006) (Fig. 12.1). In other words, improvements in the inbreds resulted in improvements in the hybrids. The genetical basis of hybrid breeding lies in the theory of inbreeding and crossbreeding.

## Inbreeding and Crossbreeding

### *Inbreeding*

We need to start with a definition of inbreeding. The following is taken from the book on population genetics by Crow and Kimura (1970). An individual with inbreeding coefficient  $f$  has a probability  $f$  that the two genes (alleles) at a particular

locus are identical by descent and a probability  $1-f$  that they are not identical, and therefore independent. If they are independent the frequencies of the genotypes will be given by the binomial formula whereas if they are identical, the frequencies of the gene pairs will be the frequencies of the alleles in the population. Thus for two alleles, 'A' and 'a' with frequencies  $p$  and  $q$ , the genotype frequencies are:

$$\begin{array}{ll} AA & p^2(1-f) + pf = p^2 + pqf \\ Aa & 2pq(1-f) = 2pq - 2pqf \\ aa & q^2(1-f) + qf = q^2 + pqf \end{array}$$

Now let us consider a population with two loci ( $A$  and  $B$ ) contributing, for example, to yield. The frequencies of the four alleles  $A$ ,  $a$ ,  $B$  and  $b$  are  $p$ ,  $q$ ,  $u$  and  $v$ , respectively, the level of inbreeding is  $f$ , and the two loci are independent and in gametic phase equilibrium. The genotype frequencies and genotypic values are given in Table 12.1. The population mean ( $M_f$ ) is:

$$\begin{aligned} M_f = & (p-q)a_A + (u-v)a_B + (p-q)(u-v)aa_{AB} + 2pq(1-f)d_A + 2uv(1-f)d_B \\ & + 2pq(u-v)(1-f)ad_{BA} + 2uv(p-q)(1-f)ad_{AB} + 4pquv(1-f)^2dd_{AB}. \end{aligned}$$

When  $f=0$  (no inbreeding), for example, an open-pollinated maize cultivar:

$$\begin{aligned} M = & (p-q)a_A + (u-v)a_B + (p-q)(u-v)aa_{AB} + 2pqd_A + 2uvd_B \\ & + 2pq(u-v)ad_{BA} + 2uv(p-q)ad_{AB} + 4pquvdd_{AB}, \text{ and} \\ M - M_f = & 2pqfd_A + 2uvfd_B + 2pq(u-v)fad_{BA} + 2uv(p-q)fad_{AB} \\ & + 4pquvf(2-f)dd_{AB} \end{aligned}$$

When  $f=1$  (complete inbreeding), the population comprises four inbred lines ( $AABB$ ,  $AAbb$ ,  $aaBB$  and  $aabb$ ) with frequencies shown in Table 12.1. The population mean is:

$$\begin{aligned} M_\infty = & (p-q)a_A + (u-v)a_B + (p-q)(u-v)aa_{AB}, \text{ and} \\ M - M_\infty = & 2pqd_A + 2uvd_B + 2pq(u-v)ad_{BA} + 2uv(p-q)ad_{AB} + 4pquvdd_{AB}. \end{aligned}$$

**Table 12.1** Genotype frequencies and genotypic values for population with two independent loci in gametic phase equilibrium: the frequencies of the four alleles  $A$ ,  $a$ ,  $B$  and  $b$  are  $p$ ,  $q$ ,  $u$  and  $v$ , respectively, and the level of inbreeding is  $f$

Genotypes	Frequencies	Genotypic value
$AABB$	$(p^2 + pqf)(u^2 + uvf)$	$a_A + a_B + aa_{AB}$
$AAbb$	$2(p^2 + pqf)uv(1-f)$	$a_A + d_B + ad_{AB}$
$AAbb$	$(p^2 + pqf)(v^2 + uvf)$	$a_A - a_B - aa_{AB}$
$AaBB$	$2pq(1-f)(u^2 + uvf)$	$a_B + d_A + ad_{BA}$
$AaBb$	$4pq(1-f)uv(1-f)$	$d_A + d_B + dd_{AB}$
$Aabb$	$2pq(1-f)(v^2 + uvf)$	$-a_B + d_A - ad_{BA}$
$aaBB$	$(q^2 + pqf)(u^2 + uvf)$	$-a_A + a_B - aa_{AB}$
$aaBb$	$2(q^2 + pqf)uv(1-f)$	$-a_A + d_B - ad_{AB}$
$aabb$	$(q^2 + pqf)(v^2 + uvf)$	$-a_A - a_B + aa_{AB}$

The fall in yield between the open-pollinated cultivar and the mean of the inbred lines depends upon the interaction of alleles at a locus (dominance) and the non-allelic interactions between loci (epistasis) which involve dominance effects, and the extent to which these affect yield in the same direction over loci and pairs of loci (directional dominance and epistasis for high yield). Unequal allele frequencies in the population also have an effect in the presence of epistasis. The fall in mean is a linear function of  $f$  in the absence of dominance by dominance epistasis ( $dd_{AB}$ ), but a function of  $f$  and  $(f)^2$  in its presence. Hallauer and Carena (2009) report typical estimates of inbreeding depression in populations of maize of 40, 10 and 10 % decreases in grain yield, plant height, and ear length, respectively, after one generation of self-pollination; and decreases of 70, 25 and 25 %, respectively after seven generations.

It is worth considering three simple cases.

(1) *Absence of epistasis:*

$$M - M_f = 2pqfd_A + 2uvfd_B \text{ and} \\ M - M_\infty = 2pqd_A + 2uvd_B$$

Hence this inbreeding depression is largest when the allele frequencies are equal (i.e., equal  $\frac{1}{2}$ , as in an  $F_2$  population) and the  $d$ 's (dominance deviations) are positive at all loci (directional dominance for high yield), giving  $\frac{1}{2}(d_A + d_B + \dots + d_N)$ . The inbreeding depression over the selfing generations is a linear function of  $f$  and is shown in Fig. 12.2 [the recurrence equation for a selfing series is  $f_n = \frac{1}{2}(1 + f_{n-1})$  where  $n$  is the generation].

(2) *Complementary epistasis for high yield* (all of the components have positive values and are of the same magnitude, and hence can be put equal to  $a_A$ ):

$$M - M_\infty = 2pqa_A + 2uva_A + 2pq(u - v)a_A + 2uv(p - q)a_A + 4pquva_A.$$

When the allele frequencies are equal (i.e., equal  $\frac{1}{2}$ , as in an  $F_2$  population), we have  $\frac{1}{2}(a_A + a_A + \frac{1}{2}a_A)$  for pairs of loci compared with  $\frac{1}{2}(a_A + a_A)$  for complete dominance and  $a_A = a_B$ .

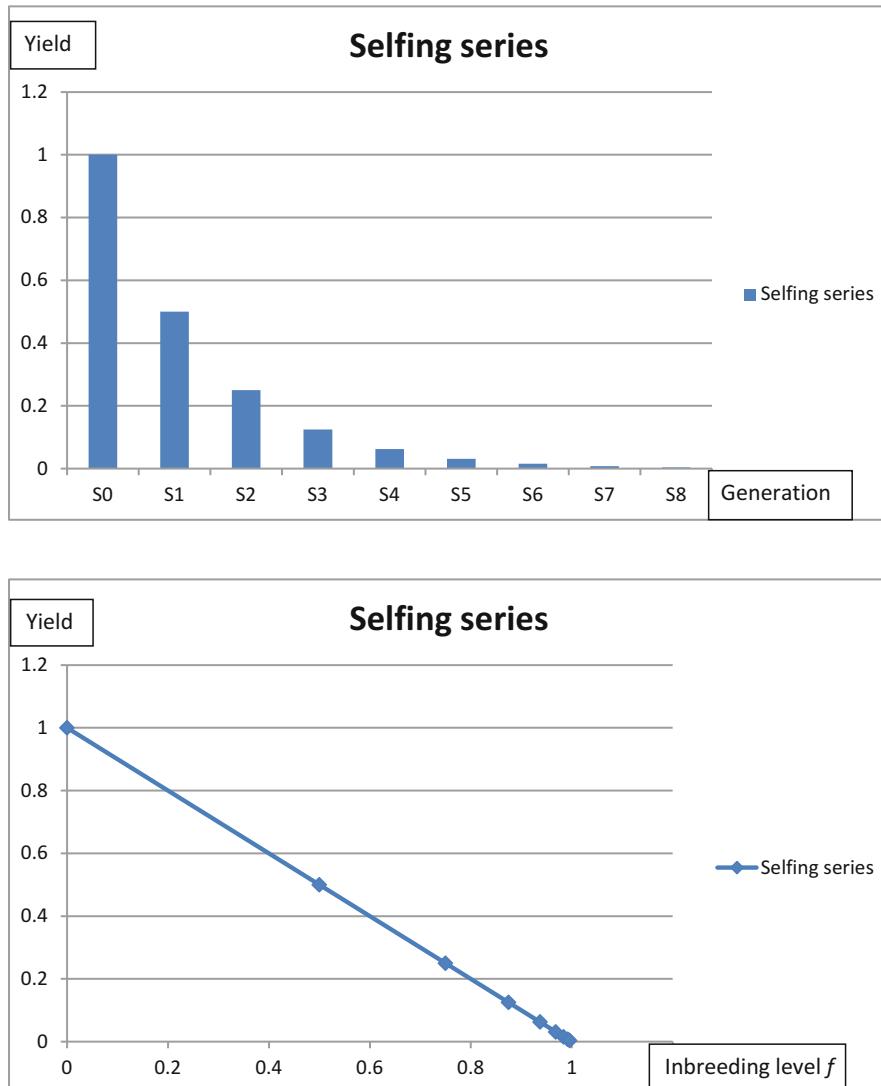
Hence the inbreeding depression is greater with complementary epistasis.

(3) *Duplicate epistasis for high yield* (all of the components are of the same magnitude, the additive and dominance effects are positive, but the interactions are negative):

$$M - M_\infty = 2pqa_A + 2uva_A - 2pq(u - v)a_A - 2uv(p - q)a_A - 4pquva_A.$$

When the allele frequencies are equal (i.e., equal  $\frac{1}{2}$ , as in an  $F_2$  population), we have  $\frac{1}{2}(a_A + a_A - \frac{1}{2}a_A)$  for pairs of loci compared with  $\frac{1}{2}(a_A + a_A)$  for complete dominance and  $a_A = a_B$ .

Hence the inbreeding depression is less with duplicate epistasis.



**Fig. 12.2** Inbreeding depression on arbitrary scale in the selfing generations from  $S_0$  with equal frequencies of two alleles and dominance for increasing effect at all loci and no epistasis

### Crossbreeding

If the inbred lines have been produced without selection (all equally viable and fertile), then crossing them in all combinations recreates the original open-pollinated cultivar (population), as shown in Box 12.1. Because we have a very large number of inbred lines, we think of line *AABB* being crossed with another line with

this genotype rather than being selfed. Hence the original cultivar can be thought of as a collection of single cross hybrids ( $F_1$ ) with the frequencies shown in Box 12.1b. We have seen that inbreeding depression for yield, for example, occurs when there is directional dominance for high yield, and that the inbreeding depression is greater when complementary epistasis occurs. The mean yield of the inbred lines increases with the frequencies of the alleles for high yield. For example, when  $p = q = u = v = \frac{1}{2}$ ,  $M_\infty$  increases from zero to  $(a_A + a_B + aa_{AB})$  as  $p$  and  $u$  increase from  $\frac{1}{2}$  to 1 (i.e., fixation of the alleles for increasing yield). The situation is more complicated for the mean of the single cross hybrids ( $M$ ). It changes from  $(\frac{1}{2}d_A + \frac{1}{2}d_B + \frac{1}{4}dd_{AB})$  to  $(a_A + a_B + aa_{AB})$ , in other words from being determined by dominance effects and their interactions to additive effects and their interactions. With classical complementary epistasis (all effects equal) and with complete dominance ( $d = a$ ,  $dd = aa = 0$ ) the mean does increase from  $1\frac{1}{4}a$  to  $3a$ , and from  $a$  to  $2a$ , respectively. In maize, we saw (Fig. 12.1) that improvements of the inbred lines have occurred and have resulted in corresponding improvements of the hybrids, which remained superior. We will return to the genetic basis of heterosis in Chap. 14. It is now time to look at the variation in the inbred lines and hybrids.

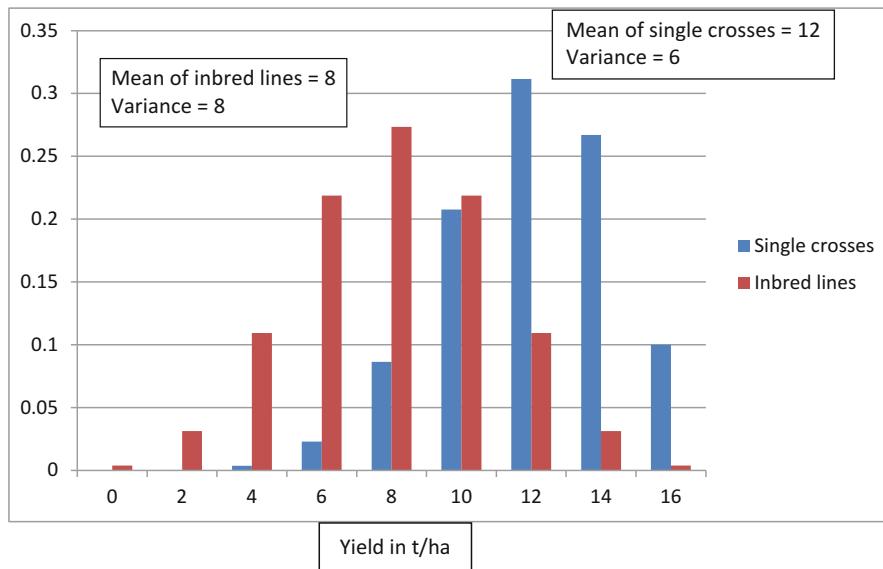
### Box 12.1

Inbred lines and population created by crossing them in all combinations

Genotypes	Frequencies	Genotypic value
<b>(a) Inbred lines</b>		
$AABB$	$pu$	$a_A + a_B + aa_{AB}$
$AAbb$	$pv$	$a_A - a_B - aa_{AB}$
$aaBB$	$qu$	$-a_A + a_B - aa_{AB}$
$aabb$	$qv$	$-a_A - a_B + aa_{AB}$
<b>(b) Population</b>		
$AABB$	$p^2u^2$	$a_A + a_B + aa_{AB}$
$AABb$	$2p^2uv$	$a_A + d_B + ad_{AB}$
$AAAb$	$p^2v^2$	$a_A - a_B - aa_{AB}$
$AaBB$	$2pqu^2$	$a_B + d_A + ad_{BA}$
$AaBb$	$4pquiv$	$d_A + d_B + dd_{AB}$
$Aabb$	$2pqv^2$	$-a_B + d_A - ad_{BA}$
$aaBB$	$q^2u^2$	$-a_A + a_B - aa_{AB}$
$aaBb$	$2q^2uv$	$-a_A + d_B - ad_{AB}$
$aabb$	$q^2v^2$	$-a_A - a_B + aa_{AB}$

### Variation Among the Inbred Lines and Single Cross Hybrids

Let us start with the simple example in Fig. 12.3 which shows the variation among the inbred lines and their hybrids for eight unlinked loci ( $n = 8$ ) of equal effect, each



**Fig. 12.3** Distribution (proportions) of single crosses (and genotypes in original open-pollinated cultivar) and inbred lines for eight unlinked loci each of equal effect, each with two alleles at equal frequencies and complete dominance for increasing effect (e.g., yield)

with two alleles at equal frequencies and complete dominance for increasing effect (e.g., yield). Thus  $a = d = 1$  at all eight loci with  $AA = Aa = 2$  and  $aa = 0$  so that the mean of all the inbred lines is 8 ( $m$ ), and their variance is 8 ( $2V_A = \sum d^2 = na^2$ ). As expected, the mean of the single crosses is 12 ( $m + \frac{1}{2}\sum d = m + \frac{1}{2}nd$ ) and their variance is 6 ( $V_A = 4$ ,  $V_D = 2$ , and  $V_A + V_D = 6$ ). So we observe inbreeding depression on selfing the open-pollinated cultivar, which comprises the population of single crosses (mean falls from 12 to 8), and then we see restoration of vigour on crossing the inbred lines to produce the population of single crosses (mean returns from 8 to 12). But how does the best inbred line compare with the best single cross? With partial or complete dominance, the best inbred (i.e., homozygous for all alleles of increasing effect) will be better than or as good as the best single cross, whereas with overdominance ( $d > a$ ) the single cross will be best. But what are the chances of finding the best inbred line, or put another way, can the larger variance of the inbred lines compensate for their lower mean? In our example with eight loci, the frequency of the best inbred is 1 in 256 compared with (approximately) 1 in 10 for the frequency of the equivalent best single crosses. In other words, the latter are over 25 times more likely to occur; and when there are 16 loci segregating, this increases to over 656 times more likely (the proportions are  $3^n/4^n$  and  $1/2^n$ , respectively, so that the ratio is  $3^n/2^n$  where  $n$  is the number of loci). We can also ask what proportion of the inbred lines and what proportion of the single crosses exceed the mean of the original open-pollinated cultivar, and hence are superior. For the inbred lines we can assume that their distribution is approximately normal with mean ' $m$ ' and variance

$2V_A$ , which in our simple model is  $na^2$ , so that the standard deviation is  $n^{1/2}a$ . The mean of the original open-pollinated cultivar is  $m + \frac{1}{2}\sum d = m + \frac{1}{2}nd$ , which is  $m + \frac{1}{2}na$  with complete dominance. Hence we are looking for the proportion of inbred lines that are  $\frac{1}{2}na/n^{1/2}a = \frac{1}{2}n^{1/2}$  standard deviations from their mean. Clearly this proportion decreases as the number of loci increases (the number of standard deviations is 1, 2, 4 and 8 for  $n = 4, 16, 64$  and 256, respectively). In contrast, the proportion of single cross hybrids that are superior to the mean of the original open-pollinated cultivar is always by definition one half. Simple arguments such as these help us to understand the success of  $F_1$  (single cross) hybrid breeding.

## F<sub>1</sub> Hybrid Maize Breeding

### *Selecting Inbred Lines to Produce the Best Possible Single Cross Hybrids*

The next question that we need to consider is how to choose the inbred lines which on hybridization will produce the best possible single cross hybrids. If we had only 32 inbred lines, we could produce all of the possible 496 single-cross hybrids and assess which is best. However, as the number of inbred lines increases, a point is reached where this becomes impossible. The world's largest seed company, Pioneer Hi-Bred, manages to evaluate a global total of about 130,000 new experimental hybrids each year (Harrington 2012). This number can be produced from all combinations of 512 inbred lines (in fact 130,816), the latter being a relatively small number. Increase this number fourfold to 2048 inbred lines and the number of single crosses is 2,096,128. Hence it is necessary to practise selection during the inbreeding process, or among the inbred lines once they have been produced, before hybrids are produced from the selected inbred lines. These hybrids are then evaluated in extensive trials to identify the superior ones for commercialization. In the days when double-cross hybrids were produced, the choice of single-cross parents had to be based on predictions as the possibilities were immense, a massive 107,880 double-cross hybrids from the 496 single-cross hybrids from 32 inbred lines (there are  $32!/(4!28!)$  combinations of 4 inbred lines drawn from 32, and for each combination there are three possible double-crosses). One commonly used prediction was the mean of the four  $F_1$ 's not used to produce the double-cross hybrid from four inbred lines (i.e., hybrids AC, AD, BC and BD for double-cross AB  $\times$  CD). As can be seen in Box 12.2, epistasis can cause a small bias in the prediction. Jenkins (1934) had originally reported a correlation of  $r = 0.75$  between observed and predicted yields of 42 double crosses. Subsequently Eberhart et al. (1964) found that the bias from epistasis was relatively small compared with experimental errors and genotype by environmental interactions. We will now examine a typical scheme for the development of inbred lines and hybrids, and its genetical basis.

**Box 12.2**

Predicting double-cross performance

Inbred lines *AABB* (A), *AAbb* (B), *aaBB* (C) and *aabb* (D)

(a) AB × CD from mean of AC, AD, BC and BD

$$AABb \times aaBb$$



*AaBB* (AC), *AaBb* (AD), *AaBb* (BC) and *Aabb* (BD) so exact prediction

(b) AC × BD from mean of AB, AD, BC and CD

$$AaBB \times Aabb$$



*AABb* (AB), *AaBb* (AD), *AaBb* (BC) and *aaBb* (CD) so exact prediction

(c) AD × BC from mean of AB, AC, BD and CD

$$AaBb \times AaBb$$



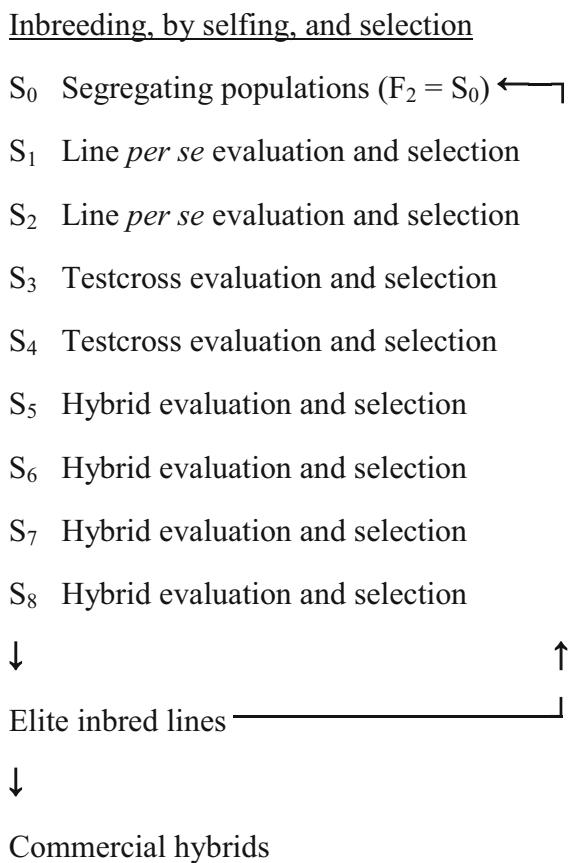
$$F_2 \text{ population with mean } \frac{1}{2}d_A + \frac{1}{2}d_B + \frac{1}{4}dd_{AB}$$

Mean of AB, AC, BD and CD (from Table 12.1) is  $\frac{1}{2}d_A + \frac{1}{2}d_B$ , so prediction out by  $\frac{1}{4}dd_{AB}$ .

### **Scheme for Development of Inbred Lines and Hybrids**

Once elite inbred lines had been developed from the best open-pollinated cultivars, further progress was primarily made by crossing these, and then crossing subsequently produced elite, inbred lines in a cyclical process (Fig. 12.4) (Betrán 2012). New desirable alleles from exotic germplasm can be introgressed into elite lines, as we shall see in Part IV, but need not concern us here. Most genetic variation is therefore created in *F*<sub>2</sub> (called *S*<sub>0</sub> in maize) populations by hybridization of inbred lines with complementary characteristics, followed by selection of progenies with the desirable traits from both parents. Inbreeding and selection can be done by any of the schemes described in Chap. 13 for inbred line cultivars, and on a similar scale. Bauman (1981) asked 130 US corn breeders how many *S*<sub>0</sub> plants they sampled from their *F*<sub>2</sub> populations to initiate inbred line development. A broad range of responses was obtained with a modal number of about 500 per population. Hallauer and Carena (2009) thought this seemed large and argued that if 10,000 nursery rows were available, the choice was samples of 500, 200 or 100 individuals from 20, 50 and 100 *F*<sub>2</sub> populations, respectively. They thought that sample sizes of 150–200 should be adequate; but larger for crosses between inbred lines of different

**Fig. 12.4** Scheme for selection of inbred lines and hybrids



origins, and smaller for closely related inbred lines. In Chap. 13 detailed arguments will be put forward in favour of a sample size of around 200, and the circumstances will be shown in which the variation between and within populations is similar, thus requiring equal resources.

There is one very important difference to consider here, namely the need to select inbred lines that combine well in a single cross hybrid rather than simply perform well in their own right. If a set of inbred lines is crossed in all combinations, the general combining ability (GCA) of a line is the mean performance of all the hybrids with this line as a parent. This can be assessed in a testcross in which the line is pollinated with a bulk of equal amounts of pollen from all of the lines. The number of testcrosses and progenies for assessment is therefore equal to the number of lines, which are also selfed. Hence testcrossing becomes feasible fairly early in the inbreeding process and selection on testcross performance replaces selection for line *per se* performance. In Bauman's (1981) survey of breeders, 60 % were starting testcrossing at S<sub>3</sub> or S<sub>4</sub>, but Hallauer and Carena (2009) reported that earlier generation (S<sub>1</sub> and S<sub>2</sub>) testcrosses are now more common.

Once the number of lines remaining is sufficiently small, they can be crossed in all combinations and the hybrids evaluated. The Pioneer Hi-Bred programme produces 130,000 new experimental hybrids each year and then reduces this number to around 15 commercial hybrids over a five-generation testing cycle. By then the commercial hybrids will have been tested at more than 150 locations and in more than 200 customers' fields (Harrington 2012). Likewise in Europe, Limagrain tested 22,774 single-cross hybrids from 1266 parental inbreds in multi-location trials over the period 1995 to 2002 (Parisseaux and Bernardo 2004).

### ***Genetic Basis of Selecting Inbreds***

The variation among the hybrids can be partitioned into that due to the general combining ability of the parents and any additional specific combining ability (SCA) arising from particular pairs of parents. The predicted performance of any hybrid ( $ij$ ) is the sum of the general combining abilities of its parents ( $i$  and  $j$ ). Any statistically significant departure from this prediction is attributed to the specific combining ability (SCA) of the two parents.

$$\text{Hybrid}_{ij} = \mu + \text{GCA}_i + \text{GCA}_j + \text{SCA}_{ij}$$

General and specific combining ability are therefore defined and estimated for a particular set of parents and their hybrids, and this should always be borne in mind (Griffing 1956). A simple example is given in Table 12.2 for a set of hybrids produced from four equally frequent types of inbred line. There are two unlinked loci, each with two alleles, and the genetic model includes dominance and epistatic effects as well as additive ones. It can be seen that the differences in general combining ability are due to additive and epistatic effects. The genotypic values of the inbred lines are perfect predictors of their general combining abilities in the absence of epistasis and still potentially good predictors in the presence of epistasis. In other words, the best hybrids do tend to come from the best inbreds. It should, however, be noted that the mean of the inbreds is less than the mean of the hybrids due to inbreeding depression in the presence of directional dominance. Specific combining abilities are due to dominance and epistatic effects. This means that in the absence of dominance and epistasis, all of the differences between the hybrids can be explained by the differences in the general combining abilities of their parents. However for traits like yield where dominance and epistasis occurs, there is no obvious predictor of specific combining abilities. In other words, once inbred lines with good general combining abilities have been selected, hybrids have to be made and assessed.

Differences in GCA can be seen in the testcross progeny of F<sub>2</sub> plants and hence form the basis of early generation selection (Table 12.2). The four homozygous genotypes have the same GCAs at F<sub>2</sub> and F<sub>∞</sub>, and the heterozygotes have values intermediate between their homozygotes. For example, AABb is intermediate between AABB and AAAb, the two homozygotes produced by continued selfing from AABb.

**Table 12.2** General (GCA) and specific (SCA) combining abilities of inbred lines (four equally frequent genotypes) and GCAs of F<sub>2</sub> population

(a) Genotypic values ( $G_{ij}$ ) of 16 types of hybrid (mean $\mu = \frac{1}{2}d_A + \frac{1}{2}d_B + \frac{1}{4}dd_{AB}$ ) and general combining abilities (GCAs) of their four inbred line parents as departures from mean of 16 hybrids ( $G_{ij} = \mu + GCA_i + GCA_j + SCA_{ij}$ )	
Genotypes	Genotypes
$\frac{1}{4}aaABBB$	$\frac{1}{4}AAAbBB$
$\frac{1}{4}per\ se$	$a_A - a_B - aa_{AB}$
$\frac{1}{4}AAABB$	$a_A + a_B + aa_{AB}$
$\frac{1}{4}AAAbb$	$a_A + d_B - aa_{AB}$
$\frac{1}{4}aaBB$	$d_A + d_B + dd_{AB}$
$\frac{1}{4}aaab$	$d_A + a_B + dd_{AB}$
$GCA$	$a_B + d_A + dd_{AB}$
(b) Specific combining abilities of pairs of parental inbred lines	
Genotypes	Genotypes
$\frac{1}{4}AAABBB$	$\frac{1}{4}AAAbBB$
$\frac{1}{4}AAAbB$	$\frac{1}{4}AAAbB$
$\frac{1}{4}aaBB$	$\frac{1}{4}aaBB$
$\frac{1}{4}aaab$	$\frac{1}{4}aaab$
(c) Genotypic values and GCAs of F <sub>2</sub> population (tester is F <sub>2</sub> ; equal frequencies of AB, Ab, aB and ab gametes), latter as departures from their mean (=population mean $\frac{1}{2}d_A + \frac{1}{2}d_B + \frac{1}{4}dd_{AB}$ )	GCAs
Genotypes	Genotypic value
1 AAbb	$a_A + a_B + aa_{AB}$
2 AAbb	$a_A + d_B + ad_{AB}$
1 AAAb	$a_A - a_B - aa_{AB}$
2 AaBB	$a_B + d_A + ad_{BA}$
4 AaBb	$d_A + d_B + dd_{AB}$
2 Aabb	$-a_B + d_A - ad_{BA}$
1 aaBB	$-a_A + a_B - aa_{AB}$
2 aaBb	$-a_A + d_B - ad_{AB}$
1 aabb	$-d_A - a_B + aa_{AB}$

## ***Inbred Line per se Versus Testcross Selection***

During the inbreeding programme it is assumed that the various types of selection practised are effective in producing hybrids which are superior to those that could be produced from a random sample of unselected inbred lines. However, some inbred line *per se* selection is required as the inbreds must have sufficient vigour and fertility to be grown as parents for F<sub>1</sub> hybrid seed production. Over the years theoretical and applied studies have been conducted to determine the relationships between inbred lines *per se* and their testcrosses, and between early and late generation testing. Maize breeders have also experimented with different kinds of testers for use in their testcrosses. It is probably more useful to think about the effectiveness of testers in achieving the ultimate goal of superior hybrids, rather than in terms of whether they are selecting for general or specific combining ability, because these terms apply to particular sets of germplasm. Interested readers can find a summary of the results in the review by Hallauer and Carena (2009). Briefly some key results are as follows. Early generation, progeny *per se* performance is not a good predictor of testcross performance. In contrast, testcross performance in early generations is a good predictor of testcross performance in later generations, as we saw in Table 12.2. For example, Jensen et al. (1983) simulated a commercial breeding programme and found that for grain yield, the correlation ( $r$ ) between S<sub>2</sub> progenies *per se* and S<sub>5</sub> testcrosses was only 0.14 whereas it was 0.67 between S<sub>2</sub> and S<sub>5</sub> testcrosses. Furthermore, Bernardo (1991) concluded that the effectiveness of early generation (generation  $n$ ) selection among testcrosses was limited by low heritability values as the genetic correlations with later generations (generation  $n'$ ) were high ( $>0.87$  from S<sub>1</sub> onwards). In fact, in the absence of selection, the testcross means of S <sub>$n$</sub>  and S <sub>$n'$</sub>  individuals or lines are identical, so that the genetic covariance between S <sub>$n$</sub>  and S <sub>$n'$</sub>  testcrosses equals the genetic variance of S <sub>$n$</sub>  testcrosses ( $V_{G(S_n)}$ ). Hence the response to testcross selection in S <sub>$n$</sub>  seen in any later generation is the heritability of the S <sub>$n$</sub>  testcrosses multiplied by the selection differential. The genetical correlation between S <sub>$n$</sub>  and S <sub>$n'$</sub>  testcrosses will be perfect ( $r = 1$ ) if S <sub>$n'$</sub>  lines from individual S <sub>$n$</sub> 's are bulked before determining the correlation. If they are kept separate (Bernardo 1991), the correlation will be  $(V_{G(S_n)}/V_{G(S_{n'})})^{1/2}$ , which equals  $[(1 + f_n)/(1 + f_{n'})]^{1/2}$  where  $f$  is the inbreeding coefficient.

These points are illustrated in Box 12.3 for an F<sub>2</sub> and a simple additive-dominance model at a single locus. For traits displaying additive genetic variation, progress with improving GCA is expected from *per se* selection, but in the presence of dominance, testcross selection is superior. With both types of selection, the response is reduced by environmental variation. Hence *per se* selection is most effective on additively inherited traits with high heritability (e.g., plant morphology, ear traits, maturity and quality characters) whereas testcross selection is more effective for traits which display directional dominance and have low heritability (e.g., grain yield). Heritabilities of testcrosses are often around 0.5 for two replicates in each of four environments but could be increased through more replications and environments. Bernardo's (1991) expected correlations between S<sub>1</sub> and

$S_\infty$  testcrosses increased from 0.67 to 0.78 as the heritability increased from 0.50 to 0.75.

### Box 12.3

Response in GCA to selection in  $F_2$  generation (*per se* and testcross)

We need to consider the regression of the testcrosses in the  $F_\infty$  generation (inbred lines) on the testcrosses and individuals in the  $F_2$  generation (individual plants from which lines derived by single seed descent).

For a single locus we have:

$F_2$ genotypes	$AA$	$Aa$	$aa$	Mean
Frequency	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	
Genotypic value	$a$	$d$	$-a$	$\frac{1}{2}d$
$F_2$ testcrosses	$\frac{1}{2}a + \frac{1}{2}d$	$\frac{1}{2}d$	$-\frac{1}{2}a + \frac{1}{2}d$	$\frac{1}{2}d$
$F_\infty$ lines from $F_2$ 's	$AA$	$\frac{1}{2}AA, \frac{1}{2}aa$	$aa$	
$F_\infty$ testcrosses ( $F_2$ bulk)	$\frac{1}{2}a + \frac{1}{2}d$	$\frac{1}{2}d$	$-\frac{1}{2}a + \frac{1}{2}d$	$\frac{1}{2}d$
$F_\infty$ testcrosses	$(\frac{1}{2}a + \frac{1}{2}d)$	$\frac{1}{2}(\frac{1}{2}a + \frac{1}{2}d)$ $\frac{1}{2}(-\frac{1}{2}a + \frac{1}{2}d)$	$(-\frac{1}{2}a + \frac{1}{2}d)$	$\frac{1}{2}d$

The tester is a population with equal frequencies of alleles  $A$  and  $a$  (i.e.,  $p = q = \frac{1}{2}$ ). Hence:

$AA$	$\times$ tester:	$\frac{1}{2}AA, \frac{1}{2}Aa;$	Mean	$\frac{1}{2}a + \frac{1}{2}d$
$Aa$	$\times$ tester:	$\frac{1}{4}AA, \frac{1}{2}Aa, \frac{1}{4}aa;$	Mean	$\frac{1}{2}d$
$aa$	$\times$ tester:	$\frac{1}{2}Aa, \frac{1}{2}aa;$	Mean	$-\frac{1}{2}a + \frac{1}{2}d$
$(\frac{1}{2}AA, \frac{1}{2}aa)$	$\times$ tester:	$\frac{1}{4}AA, \frac{1}{2}Aa, \frac{1}{4}aa;$	Mean	$\frac{1}{2}d$

Covariance with  $F_\infty$  testcrosses ( $F_2$  bulk) is same covariance with  $F_\infty$  testcrosses

$$\text{Covariance } (F_2 \text{ testcrosses}, F_\infty \text{ testcrosses}) = \text{Genetic Variance } (F_2 \text{ testcrosses}) = \frac{1}{8}a^2$$

$$\text{Genetic Variance } (F_\infty \text{ testcrosses}) = \frac{1}{4}a^2, (\text{but } \frac{1}{8}a^2 \text{ when } F_2 \text{ progeny are bulked})$$

$$\text{Covariance } (F_2, F_\infty \text{ testcrosses}) = \frac{1}{4}a^2; \text{ Genetic Variance } F_2 = \frac{1}{2}a^2 + \frac{1}{4}d^2$$

$$\text{Phenotypic Variance} = \text{Genetic Variance} + \text{Environmental Variance } (V_E)$$

$$\text{Regression of } F_\infty \text{ testcrosses on } F_2 \text{ testcrosses} = \frac{1}{8}a^2 / (\frac{1}{8}a^2 + V_{E(\text{testcrosses})})$$

$$\text{Regression of } F_\infty \text{ testcrosses on } F_2 = \frac{1}{4}a^2 / (\frac{1}{2}a^2 + \frac{1}{4}d^2 + V_{E(F2)})$$

Response to selection ( $R$ ) =  $bS = bi\sigma_P$ , where  $i$  is the intensity of selection and  $\sigma_P$  is the square root of the phenotypic variance. Hence:

(continued)

**Box 12.3 (continued)**

$$R(F_2 \text{ testcrosses}) = \frac{1}{8}a^2i / \left( \frac{1}{8}a^2 + V_{E(\text{testcrosses})} \right)^{\frac{1}{2}}$$

$$R(F_2) = \frac{1}{4}a^2i / \left( \frac{1}{2}a^2 + \frac{1}{4}d^2 + V_{E(F_2)} \right)^{\frac{1}{2}},$$

which is less than  $R(F_2 \text{ testcrosses})$  when  $(\frac{1}{4}d^2 + V_{E(F_2)}) > V_{E(\text{testcrosses})}$ .

The choice of testers to discriminate the relative combining abilities of lines was always considered important by maize breeders, but the ideal tester for all genetical situations was difficult to find in practice. However, once heterotic groups had been defined, elite inbred lines from the opposite heterotic group were used, as we will see in a later section. In Box 12.3, however, we have an  $F_2$  population so that the tester with equal frequencies of alleles 'A' and 'a' can be the  $F_2$  population or the  $F_1$  from which it was derived by selfing. With either of these testers we are selecting for general combining ability. This use of the  $F_1$  as tester is different from that in the days of double cross breeding when new inbred lines might have been sought that combined well with one of the  $F_1$  parents in the double cross.

## Hybrid Maize Seed Production

With a major crop such as maize, seed producers need to produce large quantities of high quality, genetically pure stocks to meet farmer demand. In North America alone, Pioneer offers about 300 different types of corn hybrids (Harrington 2012). The book chapter on the production of hybrid corn seed in the USA by Craig (1977) is still relevant today and worth reading for important details on the isolation of seed fields (200 m from other maize), fertility in seed fields, the control of weeds, insects, and diseases; and most importantly, roguing out off-types and checking purity of seed stocks. The inbred lines are the basic foundation seed stocks (parent seed stocks) and hence must be maintained and increased under rigid control to insure the quality of the final product. The base population for each inbred is usually bulked self-pollinated seed which adequately represents the genetic constitution of the inbred. All inbred maintenance increases are made using this base population seed, in well isolated blocks by natural random sib mating between what will be genetically identical and essentially homozygous individuals. These in turn are further multiplied to give sufficient seed of the parent seed stocks to plant and produce commercial hybrid seed. The hybrid seed is produced in isolated crossing blocks in which the female parent is detasseled or is male sterile. The most common planting pattern in seed fields is one row of pollen parent to four rows of seed parent. Split date planting of the two parents is sometimes required for them to reach the flowering stage concurrently. The correct timing of detasseling is very important to ensure genetic purity of the hybrid whilst minimizing plant damage to

maximize seed production. Detasseling can be done as a manual operation or in combination with mechanical devices, but is very expensive due to the cost of labour of thousands of workers. Since 1971 mechanical detasslers have come into widespread use in the industry, but hand detasseling is still required to remove the tassels remaining on missed plants. Pioneer, for example, still employs thousands of teenagers and adults to complete female detasseling by hand (Harrington 2012).

Not surprisingly, cheaper methods of pollen control were sought, such as cytoplasmic male sterility, genic male sterility and chemical (gametocide) pollen control. Of these, cytoplasmic male sterility has been used for maize since the 1950s, but there was a setback in 1970 as we shall see in the next section. Genic male sterility occurs when mutations in nuclear-encoded genes disrupt normal male gametogenesis. Numerous, mainly recessive, genic male sterile mutants have been identified at over 40 loci in maize (Skibbe and Schnable 2005). However, they are problematic for use in hybrid seed production because of the expense and difficulty associated with generating sterile plants and roguing fertile plants from the female rows. Attempts have been made to resolve the problem through special genetic stocks but none to date have succeeded on a commercial scale. Likewise, the use of chemical gametocides is an attractive proposition but has not yet proved feasible on a commercially worthwhile scale.

## ***Cytoplasmic Male Sterility***

Use of genetic-cytoplasmic male sterility, or CMS for short, is a three-line system, where one line (A) is the female parent of the hybrid as a result of genetic-cytoplasmic male sterility, the second line (B) is responsible for maintaining the sterility, and the third line (R) is the male parent of the hybrid that restores its fertility through a dominant restorer gene(s) (Fig. 12.5). In other words, once a source of CMS has been found, all male-fertile germplasm is either a maintainer or a restorer with respect to this source. Double cross hybrids can be produced without detasseling using the same system but only half of the plants in the farmer's field would shed pollen (Fig. 12.5). More details can be found in the book chapter by Craig (1977).

Initially the most effective male sterile cytoplasm was the one discovered in the Mexican open-pollinated cultivar Golden June at the Texas Agricultural Experimental Station; hence the name Texas (T-) cytoplasm (Rogers and Edwardson 1952). It is characterized by premature degeneration of the tapetal cell layer in developing anthers and pollen abortion. It results in complete sterility, with full fertility restored by the sporophytic expression of dominant nuclear genes *Rf1* and *Rf2a* located on chromosomes 3 and 9, respectively. By 1970 most hybrids (85 %) in the USA were produced using T-cytoplasm, but that year a southern corn leaf blight (*Bipolaris maydis*, formerly *Helminthosporium maydis*) epidemic caused crop losses of 15 %, which was more than the 12 % average expected from all diseases of corn in the United States (Tatum 1971; Ullstrup 1972). The discovery of the susceptibility of T-cytoplasm to the toxin produced by race T of southern corn leaf blight resulted in an almost complete return to detasseling in 1971 and the

Single cross hybrid

Three line system:

A-CMS; B-maintainer (inbred line); R-restorer (inbred line)

A = B nuclear genome as a result of backcrossing

A (= B) and R need to be selected for combining ability

A-CMS ( $r/rf$ ) × B (pollen) ( $rf/rf$ )



A-CMS ( $r/rf$ )    ×    R-restorer (pollen) ( $R/Rf$ )

(female fertile)    ↓

AR ( $Rf/rf$ ) (male and female fertile)

Single Cross Hybrid

Double cross hybrid

Line B is still required to maintain A-CMS; and a line E will be required to maintain D-CMS

C is an unrelated non restorer; and F is an unrelated restorer

A-CMS ( $r/rf$ ) × C (pollen) ( $rf/rf$ )    D-CMS ( $r/rf$ ) × F (pollen) ( $R/Rf$ )



AC-CMS ( $r/rf$ )    ×    DF (pollen) ( $Rf/rf$ )

(female fertile)    ↓    (male and female fertile)

Double Cross Hybrid ( $Rf/rf$  and  $rf/rf$ ) (female and half male fertile)

(A × C) × (D × F)

**Fig. 12.5** Seed production of single and double cross hybrids using genetic-cytoplasmic male sterility (CMS)

development of mechanical detasslers. Subsequently Beckett (1971) evaluated 30 maize lines which had male-sterile cytoplasm in certain inbred backgrounds. The lines could be divided into groups according to the pattern of fertility restoration observed in other backgrounds. The three main groups were Texus (T), USDA (S) and Charrua (C) cytoplasm. The S-cytoplasm was first discovered by Jenkins of USDA in a strain of Teopod maize (*Z. mexicana*) (Sofi et al. 2007). Its fertility is restored by the gametophytic expression of an allele of nuclear gene *Rf3* located on chromosome 2, so that *Rf3* pollen is viable but not *rf3* pollen. S-cytoplasm is characterized by spontaneous reversion to fertility due either to nuclear mutations or gene rearrangements, and this can be a problem. The C-cytoplasm was first discovered by Beckett (1971) in the Brazilian maize cultivar Charrua. Its fertility is restored by the sporophytic expression of dominant nuclear gene *Rf4* located on

chromosome 8. The molecular basis of these sources of cytoplasmic male sterility has been reviewed by Skibbe and Schnable (2005) and Sofi et al. (2007). The genes are mitochondrial in origin and are generated via rearrangements of the mitochondrial genome that result in novel open reading frames (mtORFs), but the details need not concern us here. The reader can consult Chen and Liu (2014) for a general review of male sterility and fertility restoration in crops.

Smith et al. (1971) found that S- and C-cytoplasm were not susceptible to race T of southern corn leaf blight and hence offered a way forward. By 1984, 8 and 3 % of US hybrids were produced using C and S cytoplasmic male sterility, respectively, but 87 % of production was with normal (N) cytoplasm and detasseling (Darrah and Zuber 1986). Today maize hybrids are produced by a combination of machine and manual detasseling as well as by using the newer forms of genetic-cytoplasmic male sterility. Although such phenotypes are associated with gene dysfunction in mitochondria, it is the chloroplasts that have emerged as ideal organs for genetic engineering of male sterility in crop plants (Sofi et al. 2007).

Since 2010, Pioneer (<http://www.pioneer.com/home/site/about/news-media/media-kits/seed-production-technology>) has been testing a new method of female parent seed production which they have named Seed Production Technology SPT. They use genetic transformation to introduce a cassette of three genes into the male-sterile female parent that converts it into a maintainer. The genes restore male-fertility, prevent the production of functional transgenic pollen and produce a colour marker which makes the seed appear pink and fluoresce in ultraviolet light. If we designate presence and absence of the cassette as  $T$  and  $t$ , on selfing the maintainer only  $t$  pollen is functional and half the progeny has genotype  $Tt$  and half  $tt$ . Optical seed sorters identify and select the  $Tt$  seed which is the maintainer. However, when the maintainer is planted alongside sterile female plants the latter are pollinated by  $t$  pollen and their progeny do not contain the cassette of transgenes; and this can be checked by scanning the seeds under ultraviolet light.

## Heterotic Groups

### Theory

A heterotic group (say 1) is germplasm that when crossed to germplasm from another heterotic group (say 2) exhibits a higher degree of heterosis than when crossed within its own group; thus  $1 \times 2$  hybrids are superior to both  $1 \times 1$  and  $2 \times 2$  hybrids. Hence on average there is an advantage in making such hybrids. We saw earlier that when a population is inbred by self-pollination without selection, the eventual conceptual outcome is a set of homozygous lines which can be intercrossed in all combinations to re-create the original open-pollinated population. So let's start with two such populations and their two sets of inbred lines which are crossed in all combinations within the two sets to create two groups (1 and 2) of single cross hybrids. For each population we have two unlinked loci ( $A$  and  $B$ ) with

**Table 12.3** Inbred lines from two populations and hybrid population created by inter-crossing the two sets of inbred lines

(a) Inbred lines			
Genotypes	Frequencies population 1	Frequencies population 2	Genotypic value
$AABB$	$pu$	$p'u'$	$a_A + a_B + aa_{AB}$
$AAbb$	$pv$	$p'v'$	$a_A - a_B - aa_{AB}$
$aaBB$	$qu$	$q'u'$	$-a_A + a_B - aa_{AB}$
$aabb$	$qv$	$q'v'$	$-a_A - a_B + aa_{AB}$

(b) Hybrid population		
Genotypes	Frequencies	Genotypic value
$AABB$	$pp'uu'$	$a_A + a_B + aa_{AB}$
$AABb$	$pp'(uv' + u'v)$	$a_A + d_B + ad_{AB}$
$AAAb$	$pp'vv'$	$a_A - a_B - aa_{AB}$
$AaBB$	$(pq' + p'q)uu'$	$a_B + d_A + ad_{BA}$
$AaBb$	$(pq' + p'q)(uv' + u'v)$	$d_A + d_B + dd_{AB}$
$Aabb$	$(pq' + p'q)vv'$	$-a_B + d_A - ad_{BA}$
$aaBB$	$qq'uu'$	$-a_A + a_B - aa_{AB}$
$aaBb$	$qq'(uv' + u'v)$	$-a_A + d_B - ad_{AB}$
$aabb$	$qq'vv'$	$-a_A - a_B + aa_{AB}$

two alleles at each ( $A, a$  and  $B, b$  where  $A$  and  $B$  have increasing effect on yield, say). However, the frequencies of the alleles are different in the two populations:  $p, q, u$  and  $v$  for one population and  $p', q', u'$  and  $v'$  for the other population. The population means are as follows.

$$M_1 = (p - q)a_A + (u - v)a_B + 2pqd_A + 2uvd_B + (p - q)(u - v)aa_{AB} \\ + 2uv(p - q)ad_{AB} + 2pq(u - v)ad_{BA} + 4pquvd_{AB}$$

$$M_2 = (p' - q')a_A + (u' - v')a_B + 2p'q'd_A + 2u'v'd_B + (p' - q')(u' - v')aa_{AB} \\ + 2u'v'(p' - q')ad_{AB} + 2p'q'(u' - v')ad_{BA} + 4p'q'u'v'dd_{AB}$$

Now let us make all of the possible hybrids between the two groups of inbred lines and calculate their mean (Table 12.3), and compare it with the average of the means of the two groups.

The results are as follows.

$$M_{12} = (pp' - qq')a_A + (uu' - vv')a_B + (pq' + p'q)d_A + (uv' + u'v)d_B \\ + (pp' - qq')(uu' - vv')aa_{AB} + (pp' - qq')(uv' + u'v)ad_{AB} \\ + (pq' + p'q)(uu' - vv')ad_{BA} + (pq' + p'q)(uv' + u'v)dd_{AB}$$

$$M_{12} - \frac{1}{2}(M_1 + M_2) = (p - p')^2d_A + (u - u')^2d_B - (p - p')(u - u')aa_{AB} \\ + (u - u')[2(pu - p'u') - (p - p') - (u - u')]ad_{AB} \\ + (p - p')[2(pu - p'u') - (p - p') - (u - u')]ad_{BA} \\ + [2(pu - p'u')(p - p' + u - u' - pu + p'u') - (p - p')(u - u')]dd_{AB}$$

We can see that the additional heterosis gets larger as the allele frequencies diverge and is at its maximum when different alleles are fixed in the two populations. It also increases with the extent of directional dominance and epistasis for increased trait value (e.g., yield). With our simple model the two extremes are as follows.

With  $p = u = 1$  and  $p' = u' = 0$ :

$$\begin{aligned} M_1 &= a_A + a_B + aa_{AB} \\ M_2 &= -a_A - a_B + aa_{AB} \\ M_{12} &= d_A + d_B + dd_{AB} \\ M_{12} - \frac{1}{2}(M_1 + M_2) &= d_A + d_B - aa_{AB} + dd_{AB} \end{aligned}$$

With  $p = u' = 1$  and  $p' = u = 0$ :

$$\begin{aligned} M_1 &= a_A - a_B - aa_{AB} \\ M_2 &= -a_A + a_B - aa_{AB} \\ M_{12} &= d_A + d_B + dd_{AB} \\ M_{12} - \frac{1}{2}(M_1 + M_2) &= d_A + d_B + aa_{AB} + dd_{AB} \end{aligned}$$

In the absence of epistasis we have the simple result (Falconer and Mackay 1996):

$$M_{12} - \frac{1}{2}(M_1 + M_2) = \Sigma(p_k - p_k')^2 d_k \text{ where summation is over } n \text{ loci} \\ (k = A \text{ to } N, \text{say}).$$

It is important to appreciate that here we have measured heterosis from the mid-parent (mid-population) value. In Chap. 14 we will need to look at the causes of heterosis and consider better parent (population) heterosis. Before looking at heterotic groups in maize, it is worth looking at the circumstances where the simple result given above can be used to predict heterosis.

### **Predicting Heterosis**

If we assume that heterosis is due to dominance, and that the dominance deviation ( $d_k$ ) at a locus is independent of the squared difference in allele frequency between parental populations (lines) at that locus ( $(p_k - p_k')^2$ ), then the expected value ( $E$ ) of  $(p_k - p_k')^2 d_k$  is the expected value of  $(p_k - p_k')^2$  multiplied by the expected value of  $d_k$  (Bulmer 1967).

Hence  $(1/n)\Sigma(p_k - p_k')^2 d_k = E(p_k - p_k')^2 Ed_k$  and

$$\Sigma(p_k - p_k')^2 d_k = nE(p_k - p_k')^2 Ed_k$$

as pointed out by Amuzu-Aweh et al. (2013).

Furthermore, the mean of the hybrid population is:

$$M_{12} = \frac{1}{2}(M_1 + M_2) + \Sigma \left( p_k - p_k' \right)^2 d_k = \frac{1}{2}(M_1 + M_2) + nE \left( p_k - p_k' \right)^2 Ed_k$$

Now if we have a set of  $m$  populations or inbred lines, the mean of the hybrid between population (line)  $i$  and population (line)  $j$  will be:

$$\begin{aligned} M_{ij} &= \frac{1}{2}(M_i + M_j) + \Sigma (p_{ik} - p_{jk})^2 d_k \\ &= \frac{1}{2}(M_i + M_j) + nE(p_{ik} - p_{jk})^2 Ed_k, \text{ summation is over } k \text{ loci.} \end{aligned}$$

For the  $m$  populations (lines),  $nEd_k$  (the number of loci at which at least two lines differ multiplied by the expected value of the dominance deviations at those loci) will be an unknown constant. Hence the variation in  $M_{ij}$  will result from variation in  $M_i$  and  $M_j$ , and variation in  $E(p_{ik} - p_{jk})^2$ , the latter being variation in the squared difference in allele frequency between parental populations (lines)  $i$  and  $j$ . If a random sample (over loci) of molecular markers such as SNPs reflects the variation at the  $k$  loci in the  $m$  populations (lines), then it is possible to obtain an estimate of  $E(p_{ik} - p_{jk})^2$ . It is then possible to predict the hybrid with the most heterosis and to select parental populations or lines for crossing on the basis of their midparental value and predicted relative heterosis. The actual heterosis will depend upon  $nEd_k$ , which equals  $\Sigma d_k$ , and hence will be greatest with large  $d$ 's of the same sign (directional dominance). An example is given in Box 12.4. Whilst  $nEd_k E(p_{ik} - p_{jk})^2$  is an unbiased estimator of  $\Sigma (p_{ik} - p_{jk})^2 d_k$ , in any particular hybrid, overestimates and underestimates can occur by chance. The example also brings out the point that one needs to know the value of  $nEd_k$  in order to make a reasonable prediction of the best hybrid, but this is only available after making crosses from the regression of observed heterosis on  $E(p_{ik} - p_{jk})^2$ . Balestre et al. (2009) evaluated the potential of genetic distances estimated by microsatellite markers for the prediction of the performance of single-cross maize hybrids. They found that genetic distances had moderate predictive ability for grain yield (0.546), specific combining ability (0.567) and heterosis (0.661).

#### Box 12.4

Predicting heterosis from the squared difference in allele frequency (SDAF)

Consider a cross between two inbred lines that differ at eight unlinked loci ( $A$  to  $H$ ) of equal additive effect ( $a = 1$ ) but whose dominance deviations follow a binomial distribution with mean (expected value) 0.5:

$$d_A = 0.2, d_B = d_C = d_D = 0.4, d_E = d_F = d_G = 0.6 \text{ and } d_H = 0.8. \\ (nEd_k = 8 \times 0.5 = 4 = \Sigma d_k)$$

Then produce the complete set of 256 unique recombinant inbred lines from the cross and let the mean of all the lines be 10:

(continued)

**Box 12.4** (continued)

1	<i>aabbccddeeffgghh</i>	Mean = 2
8	<i>AAbbccddeeffgghh</i> etc.	Mean = 4
28	<i>AABBccddeeffgghh</i> etc.	Mean = 6
56	<i>AABBCCDdeeffgghh</i> etc.	Mean = 8
70	<i>AABBCCDDeeffgghh</i> etc.	Mean = 10
56	<i>AABBCCDDEEfFgghh</i> etc.	Mean = 12
28	<i>AABBCCDDEEFFgghh</i> etc.	Mean = 14
8	<i>AABBCCDDEEFFGGhh</i> etc.	Mean = 16
1	<i>AABBCCDDEEFFGGHH</i>	Mean = 18

**Now select some pairs of parents:**

*Example 1*

$$AABBCCDDEEFFGGHH \times aabbccddeeffgghh \rightarrow AaBbCcDdEeFfGgHh$$

$$M_i = 18 \quad M_j = 2 \quad M_{ij} = 14$$

$$\frac{1}{2}(M_i + M_j) = 10; E(pi_k - pj_k)^2 = (1/n)\Sigma(pi_k - pj_k)^2 = 8/8 = 1; nEd_k = 4$$

$$\text{Predicted } M_{ij} = \frac{1}{2}(M_i + M_j) + nEd_k E(pi_k - pj_k)^2 = 10 + 4 \times 1 = 14$$

Prediction correct when parents differ at all eight loci.

*Example 2*

$$AABBCCDDeeffggHH \times AAAbccddEEFFGGHH \rightarrow AABbCcDdEeFfGgHH$$

$$M_i = 12 \quad M_j = 12 \quad M_{ij} = 15$$

$$\frac{1}{2}(M_i + M_j) = 12; E(pi_k - pj_k)^2 = (1/n)\Sigma(pi_k - pj_k)^2 = 6/8 = 0.75; nEd_k = 4$$

$$\text{Predicted } M_{ij} = \frac{1}{2}(M_i + M_j) + nEd_k E(pi_k - pj_k)^2 = 12 + 4 \times 0.75 = 15$$

Prediction correct because  $nEd_k E(pi_k - pj_k)^2$  equals  $\Sigma(pi_k - pj_k)^2 d_k$  (which is  $0 \times 0.2 + 1 \times 0.4 + 1 \times 0.4 + 1 \times 0.4 + 1 \times 0.6 + 1 \times 0.6 + 1 \times 0.6 + 0 \times 0.8 = 3$ )

*Example 3*

$$AABBCCDDEEffgghh \times AABbccddeFFGGHH \rightarrow AABCcDdEeFfGgHh$$

$$M_i = 12 \quad M_j = 12 \quad M_{ij} = 15.4$$

$$\frac{1}{2}(M_i + M_j) = 12; E(pi_k - pj_k)^2 = (1/n)\Sigma(pi_k - pj_k)^2 = 6/8 = 0.75; nEd_k = 4$$

Predicted

$$M_{ij} = \frac{1}{2}(M_i + M_j) + nEd_k E(pi_k - pj_k)^2 = 12 + 4 \times 0.75 = 15$$

(continued)

**Box 12.4** (continued)

Prediction no longer correct because  $nEd_kE(pi_k - pj_k)^2$  does not equal  $\Sigma(pi_k - pj_k)^2d_k$  (which is  $0 \times 0.2 + 0 \times 0.4 + 1 \times 0.4 + 1 \times 0.4 + 1 \times 0.6 + 1 \times 0.6 + 1 \times 0.6 + 1 \times 0.8 = 3.4$ )

*Example 4*

$AABBCCddeeFFGGHH \times aabbccDDEEFFGGHH \rightarrow AaBbCcDdEeFfGGHH$   
 $M_i = 12 \quad M_j = 12 \quad M_{ij} = 14.6$   
 $\frac{1}{2}(M_i + M_j) = 12; E(pi_k - pj_k)^2 = (1/n)\Sigma(pi_k - pj_k)^2 = 6/8 = 0.75; nEd_k = 4$   
Predicted  $M_{ij} = \frac{1}{2}(M_i + M_j) + nEd_kE(pi_k - pj_k)^2 = 12 + 4 \times 0.75 = 15$

Prediction no longer correct because  $nEd_kE(pi_k - pj_k)^2$  does not equal  $\Sigma(pi_k - pj_k)^2d_k$  (which is  $1 \times 0.2 + 1 \times 0.4 + 1 \times 0.4 + 1 \times 0.4 + 1 \times 0.6 + 1 \times 0.6 + 0 \times 0.6 + 0 \times 0.8 = 2.6$ ) This time prediction is an overestimate by 0.4 compared with an underestimate of 0.4 in the previous example. Averaged over all possible hybrids the mean bias is zero. In other words  $nEd_kE(pi_k - pj_k)^2$  is an unbiased estimator of  $\Sigma(pi_k - pj_k)^2d_k$ .

*Example 5*

$AABBCCDDEeffGGHH \times AABBCCDDeFFGGHH \rightarrow AABBCCDDeFfGGHH$   
 $M_i = 16 \quad M_j = 16 \quad M_{ij} = 17.2$   
 $\frac{1}{2}(M_i + M_j) = 16; E(pi_k - pj_k)^2 = (1/n)\Sigma(pi_k - pj_k)^2 = 2/8 = 0.25; nEd_k = 4$   
Predicted  $M_{ij} = \frac{1}{2}(M_i + M_j) + nEd_kE(pi_k - pj_k)^2 = 16 + 4 \times 0.25 = 17$

Prediction no longer correct because  $nEd_kE(pi_k - pj_k)^2$  does not equal  $\Sigma(pi_k - pj_k)^2d_k$  (which is  $0 \times 0.2 + 0 \times 0.4 + 0 \times 0.4 + 0 \times 0.4 + 1 \times 0.6 + 1 \times 0.6 + 0 \times 0.6 + 0 \times 0.8 = 1.2$ ) This is an example of what happens when the best inbred lines are selected before making any crosses. The cross has the highest midparental value, but lowest heterosis out of the few crosses we have considered.

## Heterotic Groups in Maize

Heterotic groups in maize were established empirically. Two of the best known heterotic groups of the US Corn Belt are derived from the open-pollinated cultivars Reid Yellow Dent from central Illinois and Lancaster Sure Crop from southeastern Pennsylvania. Because of their origins, Hallauer and Carena (2009) thought that they would differ in their allele frequencies at many loci, and that this was the explanation

of the heterosis seen in their hybrids. One could envisage different alleles being fixed so that the two types of maize were true breeding for their typical characteristics, with heterozygosity in the hybrids resulting in hybrid vigour. Later assignment of inbred lines to heterotic groups was primarily based on whether they were of Reid Yellow Dent origin or not; but these evolved to be designated Iowa Stiff Stalk Synthetic and non-Iowa Stiff Stalk Synthetics. A third important heterotic group is Iodent (van Heerwaarden et al. 2012). The recent results of van Heerwaarden et al. (2012) offer a different explanation of the origin of heterotic groups. They used 46,000 SNPs to analyze 99 classic North America landraces, 94 inbreds from before the 1950s, 70 advanced public lines from the 1960s and 70s and 137 elite commercial lines from the 1980s and 90s. They found that the three modern heterotic groups (Iowa Stiff Stalk Synthetic (SS), Non-Stiff Stalk (NSS) and Iodent (IDT)) were the result of ongoing divergence from a relatively homogeneous landrace population, with Yellow Dent as the main ancestral contributor to all three heterotic groups in the elite commercial lines. Nevertheless, small differences in contribution of Yellow Dent and Lancaster to SS and NSS lent some justification to the traditional distinction between these two heterotic groups on the basis of their landrace ancestry.

Today breeding plans emphasize recycling of inbred lines within heterotic groups in order to maintain favourable genetic combinations that enhance heterosis. Selection during inbreeding within each heterotic group is for combining ability with the other heterotic group, using a sample of bulked pollen or the best inbreds from that group. It can be seen in Table 12.4 that combining ability depends on allele frequencies in the tester population with contributions from dominance and epistatic effects as well as from additive effects. Variation in combining ability depends on the genotype frequencies in the population being assessed. Hence it is better to talk about combining ability with the other population rather than thinking in terms of selecting for general and specific combining ability.

**Table 12.4** Combining abilities of genotypes in one initial population (one heterotic group) with bulk of gametes from other population (other heterotic group) with gamete frequencies for  $AB$ ,  $Ab$ ,  $aB$  and  $ab$  of  $pu$ ,  $pv$ ,  $qu$  and  $qv$

Genotypes	Genotypic value	Combining ability
$AABB$	$a_A + a_B + aa_{AB}$	$pa_A + ua_B + qd_A + vd_B + pu aa_{AB} + pv ad_{AB} + quad_{BA} + qv dd_{AB}$
$AABb$	$a_A + d_B + ad_{AB}$	$pa_A + \frac{1}{2}(u - v)a_B + qd_A + \frac{1}{2}d_B + \frac{1}{2}p(u - v)aa_{AB} + \frac{1}{2}pd_{AB} + \frac{1}{2}q(u - v)ad_{BA} + \frac{1}{2}qdd_{AB}$
$AAbb$	$a_A - a_B - aa_{AB}$	$pa_A - va_B + qd_A + ud_B - pv aa_{AB} + pu ad_{AB} - qv ad_{BA} + qu dd_{AB}$
$AaBB$	$a_B + d_A + ad_{BA}$	$\frac{1}{2}(p - q)a_A + ua_B + \frac{1}{2}d_A + vd_B + \frac{1}{2}(p - q)uaa_{AB} + \frac{1}{2}(p - q)vad_{AB} + \frac{1}{2}uad_{BA} + \frac{1}{2}vdd_{AB}$
$AaBb$	$d_A + d_B + dd_{AB}$	$\frac{1}{2}(p - q)a_A + \frac{1}{2}(u - v)a_B + \frac{1}{2}d_A + \frac{1}{2}d_B + \frac{1}{4}(p - q)(u - v)aa_{AB} + \frac{1}{4}(p - q)ad_{AB} + \frac{1}{4}(u - v)ad_{BA} + \frac{1}{4}dd_{AB}$
$Aabb$	$-a_B + d_A - ad_{BA}$	$\frac{1}{2}(p - q)a_A - va_B + \frac{1}{2}d_A + ud_B + \frac{1}{2}(-p + q)va a_{AB} + \frac{1}{2}(p - q)uad_{AB} - \frac{1}{2}vad_{BA} + \frac{1}{2}udd_{AB}$
$aaBB$	$-a_A + a_B - aa_{AB}$	$-qa_A + ua_B + pd_A + vd_B - qua a_{AB} - qv ad_{AB} + pu ad_{BA} + pv dd_{AB}$
$aaBb$	$-a_A + d_B - ad_{AB}$	$-qa_A + \frac{1}{2}(u - v)a_B + pd_A + \frac{1}{2}d_B + \frac{1}{2}q(-u + v)aa_{AB} - \frac{1}{2}qad_{AB} + \frac{1}{2}p(u - v)ad_{BA} + \frac{1}{2}pdd_{AB}$
$aabb$	$-a_A - a_B + aa_{AB}$	$-qa_A - va_B + pd_A + ud_B + qva a_{AB} - quad_{AB} - pva d_{BA} + pudd_{AB}$

## Reciprocal Recurrent Selection

### Theory

The establishment of heterotic groups saw the development of selection schemes designed to improve the cross between two populations from different groups, rather than the populations *per se*. The same is true where the use of genetic-cytoplasmic male sterility divides the germplasm into maintainer and restorer populations. The schemes are known as reciprocal recurrent selection (Comstock et al. 1949; Hallauer and Eberhart 1970). Selection of parents (usually their selfed progeny) for the next cycle of crossing within one population is based on their combining ability with the other population, and the response to selection of interest is in the cross of the two populations. The advantage of the method over those that select independently within each population occurs for those loci that display overdominance ( $d > a$ ) or pseudo-overdominance from tight repulsion linkage ( $Ab/aB$ ) of alleles of increasing effect. This can be seen in Box 12.5 and Fig. 12.6 with a simple additive-dominance model (i.e., no epistasis). When the frequencies of alleles 'A' and 'B' are high in the tester population, genotypes  $aabb$ ,  $aaBb$  and  $Aabb$  have the highest combining abilities and are selected (i.e., alleles 'a' and 'b' increase in frequency). In contrast, when the frequencies of alleles 'a' and 'b' are high in the tester population, there are big differences in combining ability and genotypes  $AABB$ ,  $AAbb$  and  $AaBB$  are selected (i.e., alleles 'A' and 'B' increase in frequency). Although in our particular model there are no differences in combining ability when  $p = u = \frac{3}{4}$  ( $q = v = \frac{1}{4}$ ), this is an unstable equilibrium in the presence of the chance fluctuations of allele frequencies that will always occur. Hence in Fig. 12.6, the two eventual outcomes are fixation of 'A' in one population and 'a' in the other, and likewise and independently (if unlinked) for 'B' and 'b' (cross mean 4.0). Independent selection within each population would result in a stable equilibrium at  $p = u = \frac{3}{4}$  and  $p' = u' = \frac{3}{4}$ . At those loci which do not display overdominance (or pseudo-overdominance), the alleles of increasing effect would eventually be fixed in both populations, and this is the explanation for the continuing improvement of inbred lines of maize.

### Practice

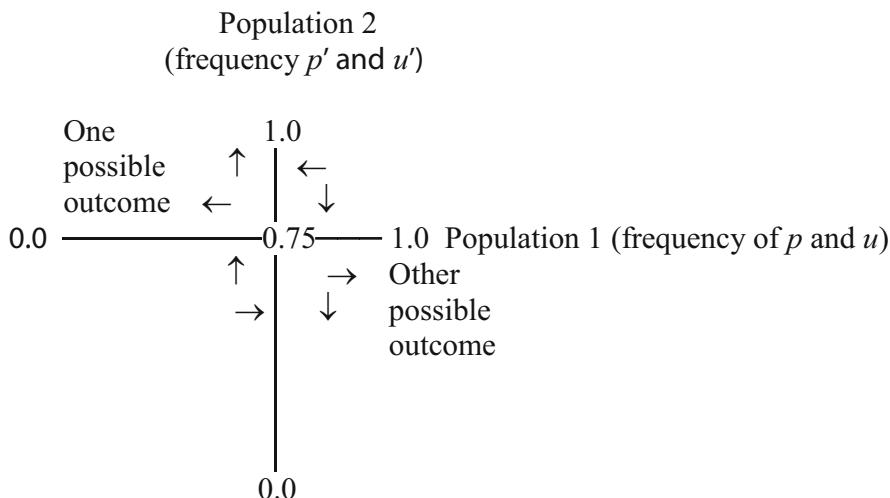
With maize plants that produce only one ear, these are used to produce the selfed seed and the pollen of each plant is used in matings to several females of the other population. The seed from these families is bulked to give a half-sib family for evaluation. However, with maize plants that have two fertile ears, one can be selfed and the other pollinated with a bulk of pollen from the other population to produce the half-sib family. Furthermore, in this situation, it is possible to practice full-sib reciprocal recurrent selection. Jones et al. (1971) used computer simulations to demonstrate the superiority of full-sib over half-sib selection for a number of

genetical models. Hallauer and Carena (2009) summarized the results for grain yield of 20 published programmes with 1–11 cycles of selection: the average response per cycle was 4.8 % (range 0.8–7.5 %), and in five programmes where data were available, midparent heterosis increased on average from 7.3 to 37.4 % over 6–11 cycles of selection.

### Box 12.5

Combining abilities of genotypes in population in table below with another population with allele frequencies for  $A$ ,  $a$ ,  $B$  and  $b$  of  $p$ ,  $q$ ,  $u$  and  $v$ ; simple additive-dominance model with overdominance at each locus:  $d_A = d_B = 2a_A = 2a_B = 2$

Genotypes	Genotypic value	Combining ability
(a) $p = u = \frac{1}{8}$ , $q = v = \frac{1}{8}$		
$AABB$	$a_A + a_B = 2.0$	$\frac{1}{8}a_A + \frac{1}{8}a_B + \frac{1}{8}d_A + \frac{1}{8}d_B = 2.25$
$AABb$	$a_A + d_B = 3.0$	$\frac{1}{8}a_A + \frac{1}{8}a_B + \frac{1}{8}d_A + \frac{1}{2}d_B = 2.50$
$AAAb$	$a_A - a_B = 0.0$	$\frac{1}{8}a_A - \frac{1}{8}a_B + \frac{1}{8}d_A + \frac{1}{8}d_B = 2.75$
$AaBB$	$a_B + d_A = 3.0$	$\frac{1}{8}a_A + \frac{1}{8}a_B + \frac{1}{2}d_A + \frac{1}{8}d_B = 2.50$
$AaBb$	$d_A + d_B = 4.0$	$\frac{1}{8}a_A + \frac{1}{8}a_B + \frac{1}{2}d_A + \frac{1}{2}d_B = 2.75$
$Aabb$	$-a_B + d_A = 1.0$	$\frac{1}{8}a_A - \frac{1}{8}a_B + \frac{1}{2}d_A + \frac{1}{8}d_B = 3.00$
$aaBB$	$-a_A + a_B = 0.0$	$-\frac{1}{8}a_A + \frac{1}{8}a_B + \frac{1}{8}d_A + \frac{1}{8}d_B = 2.75$
$aaBb$	$-a_A + d_B = 1.0$	$-\frac{1}{8}a_A + \frac{1}{8}a_B + \frac{1}{8}d_A + \frac{1}{2}d_B = 3.00$
$aabb$	$-a_A - a_B = -2.0$	$-\frac{1}{8}a_A - \frac{1}{8}a_B + \frac{1}{8}d_A + \frac{1}{8}d_B = 3.25$
(b) $p = u = \frac{1}{4}$ , $q = v = \frac{1}{4}$		
$AABB$	$a_A + a_B = 2.0$	$\frac{1}{4}a_A + \frac{1}{4}a_B + \frac{1}{4}d_A + \frac{1}{4}d_B = 2.5$
$AABb$	$a_A + d_B = 3.0$	$\frac{1}{4}a_A + \frac{1}{4}a_B + \frac{1}{4}d_A + \frac{1}{2}d_B = 2.5$
$AAAb$	$a_A - a_B = 0.0$	$\frac{1}{4}a_A - \frac{1}{4}a_B + \frac{1}{4}d_A + \frac{1}{4}d_B = 2.5$
$AaBB$	$a_B + d_A = 3.0$	$\frac{1}{4}a_A + \frac{1}{4}a_B + \frac{1}{2}d_A + \frac{1}{4}d_B = 2.5$
$AaBb$	$d_A + d_B = 4.0$	$\frac{1}{4}a_A + \frac{1}{4}a_B + \frac{1}{2}d_A + \frac{1}{2}d_B = 2.5$
$Aabb$	$-a_B + d_A = 1.0$	$\frac{1}{4}a_A - \frac{1}{4}a_B + \frac{1}{2}d_A + \frac{1}{4}d_B = 2.5$
$aaBB$	$-a_A + a_B = 0.0$	$-\frac{1}{4}a_A + \frac{1}{4}a_B + \frac{1}{4}d_A + \frac{1}{4}d_B = 2.5$
$aaBb$	$-a_A + d_B = 1.0$	$-\frac{1}{4}a_A + \frac{1}{4}a_B + \frac{1}{4}d_A + \frac{1}{2}d_B = 2.5$
$aabb$	$-a_A - a_B = -2.0$	$-\frac{1}{4}a_A - \frac{1}{4}a_B + \frac{1}{4}d_A + \frac{1}{4}d_B = 2.5$
(c) $p = u = \frac{1}{8}$ , $q = v = \frac{7}{8}$		
$AABB$	$a_A + a_B = 2.0$	$\frac{1}{8}a_A + \frac{1}{8}a_B + \frac{1}{8}d_A + \frac{1}{8}d_B = 3.75$
$AABb$	$a_A + d_B = 3.0$	$\frac{1}{8}a_A - \frac{1}{8}a_B + \frac{1}{8}d_A + \frac{1}{2}d_B = 2.50$
$AAAb$	$a_A - a_B = 0.0$	$\frac{1}{8}a_A - \frac{1}{8}a_B + \frac{1}{8}d_A + \frac{1}{8}d_B = 1.25$
$AaBB$	$a_B + d_A = 3.0$	$-\frac{1}{8}a_A + \frac{1}{8}a_B + \frac{1}{2}d_A + \frac{1}{8}d_B = 2.50$
$AaBb$	$d_A + d_B = 4.0$	$-\frac{1}{8}a_A - \frac{1}{8}a_B + \frac{1}{2}d_A + \frac{1}{2}d_B = 1.25$
$Aabb$	$-a_B + d_A = 1.0$	$-\frac{1}{8}a_A - \frac{1}{8}a_B + \frac{1}{2}d_A + \frac{1}{8}d_B = 0.00$
$aaBB$	$-a_A + a_B = 0.0$	$-\frac{7}{8}a_A + \frac{1}{8}a_B + \frac{1}{8}d_A + \frac{1}{8}d_B = 1.25$
$aaBb$	$-a_A + d_B = 1.0$	$-\frac{7}{8}a_A - \frac{1}{8}a_B + \frac{1}{8}d_A + \frac{1}{2}d_B = 0.00$
$aabb$	$-a_A - a_B = -2.0$	$-\frac{7}{8}a_A - \frac{1}{8}a_B + \frac{1}{8}d_A + \frac{1}{8}d_B = -1.25$
(d) $p = u = 0$ , $q = v = 1$ ; $AABB$ combining ability with $aabb = 4.00$		



**Fig. 12.6** Half-sib reciprocal recurrent selection in two populations each with same two alleles at same two loci,  $A$ ,  $a$  and  $B$ ,  $b$ ; frequencies of  $A$  and  $B$  in two populations are  $p$ ,  $u$  and  $p'$ ,  $u'$ , respectively; simple additive-dominance model with overdominance at each locus:  $d_A = d_B = 2a_A = 2a_B = 2$

## Botanical Varieties of *Brassica oleracea*

Human selection in *Brassica oleracea* has produced a number of morphologically distinct but interfertile vegetable and forage crops: broccoli, Brussels sprouts, cabbage, cauliflower, Chinese kale, curled kale, kohlrabi, marrow-stem kale and thousand-head kale. They can be regarded as heterotic groups for breeding forage crops as Watts (1970) found that their yields were surpassed by the hybrids between them. Two contrasting approaches have been taken to breeding forage crops from this germplasm that raise the question of how to combine genes from heterotic groups. In one approach, hybrids were made between the various types of *B. oleracea* in order to find the best combination (Thompson 1967; Watts 1970). Then inbred lines of the two chosen types were selected and used to produce hybrids. Cultivar Bittern was bred at the former Plant Breeding Institute in Cambridge, England, and is a hybrid between marrow-stem kale and Brussels sprouts. It combines the high dry-matter yield of marrow-stem kale with the winter hardiness of Brussels sprouts (Thompson 1967). Morphologically it looks like a marrow-stem kale with the leaves of Brussels sprouts. In the other approach half-sib family selection for yield of digestible organic matter (DOM) was practised on a population which initially included thousand-head kales, curled kales, Brussels sprouts and cabbages, in addition to marrow-stem kales (Bradshaw and Mackay 1985). No selection was practised for morphological type, but after four generations of selection for DOM yield the population consisted entirely of medium to tall marrow-stem kales, but with sufficient variation in leaf

morphology to create uniformity problems in testing half-sib families for Distinctness, Uniformity and Stability. Subsequently, for greater uniformity and ease of production of relatively cheap seed, open-pollinated cultivars Caledonian and Grampian were bred from a population of marrow-stem kales from different parts of the world (Bradshaw and Wilson 2012). However, in the longer term the population of botanical varieties could have been used as a source of inbred lines for hybrid cultivars, and this leaves open the question of the best long term strategy for combining genes from heterotic groups of maize and other crops. We are now going to look at hybrid breeding in crops other than maize, starting with kale.

## Kale and Vegetables of *Brassica oleracea*

### Kale

Kale was an important fodder crop in Great Britain in the 1950s and 1960s, having been introduced from France during the early 1900s. Breeding work started at the former Plant Breeding Institute in Cambridge in 1951 with the aim of replacing the old open-pollinated cultivars with more uniform and higher yielding double-cross hybrids (Thompson 1967). Maris Kestrel was released in 1967 and proved a very successful cultivar which is still grown today. However, in order to produce enough seed to meet demand, Thompson (1964) devised the triple-cross method of production for Maris Kestrel, and subsequent hybrids such as cultivar Bittern, based on an understanding of self-incompatibility in *Brassica oleracea*. Thompson (1957) demonstrated that self-incompatibility was due to a sporophytic system with the following features. There is a multiple series of *S* alleles at a single locus, with over 30 different *S*-allele homozygotes produced by Thompson (1967). Dominance of one allele over another may be found in the pollen or in the style, or the two alleles may be active in both pollen and style. Thus the four combinations for alleles *S<sub>a</sub>* and *S<sub>b</sub>* are *S<sub>a</sub>* dominant over *S<sub>b</sub>* in both pollen and stigma, *S<sub>a</sub>* dominant in the pollen but both alleles active in the stigma, *S<sub>a</sub>* dominant in the stigma but both alleles active in the pollen, and finally, both alleles active in both pollen and stigma.

The principles of the scheme for triple-cross production are shown in Fig. 12.7. Readers can learn more about how the practical details evolved from the account by Thompson (1967). Inbred lines can be produced and maintained by pollinating a bud with pollen from another flower on the same plant at least 2 days before the bud opens. Lines homozygous for different *S* alleles need to be recognized. The production of 100 % triple-cross hybrid seed by insect (bee) pollination depends on the dominance relationships between *S* alleles in the constituent inbred lines. Alleles 3 and 6 must be active in both pollen and stigma in the presence of alleles 1 and 2, and 4 and 5, respectively, to prevent any self and cross-pollination within the three-way crosses. This ensures only cross-pollination between them in producing certified seed of the final triple-cross hybrid which is sold to farmers. In practice dominance was not

INBRED LINES	A	B	C	D	E	F	Self in bud in glasshouse
S ALLELES	<i>11</i>	22	33	44	55	66	
SINGLE CROSSES	A × B			D × E			Small isolation plots
		↓			↓		Insect (bee) pollination
		AB			DE		
		<i>12</i>			<i>45</i>		
THREE-WAY CROSSES	AB	×	C	DE	×	F	Field multiplications
	2	↓	1	2	↓	1	Drilled in these ratios
	AC	BC		DF	EF		
	<i>13</i>	<i>23</i>		<i>46</i>	<i>56</i>		Bold is active S allele
FINAL TRIPLE-CROSS	(AC BC)	×	(DF EF)				Field multiplication

**Fig. 12.7** Production of triple-cross hybrid kale

a serious restriction on the arrangements of inbreds as it was very rare between 23 out of 28 S alleles tested (Thompson 1967). Self-compatible inbreds and self-incompatible ones segregating for self-compatibility do occur and are unsuitable for hybrid seed production. Thompson and Taylor (1967) found higher frequencies of recessive S alleles and associated self-compatibility in stocks of many other botanical varieties of *B. oleracea*, with implications for the use of self-incompatibility in F<sub>1</sub> hybrid vegetable production. It is important to use S alleles high in the dominance series in order to produce highly-self incompatible inbred lines for almost 100 % F<sub>1</sub> hybrid seed. Today the incompatibility system is well characterized at the molecular level in terms of two genes encoding two distinct proteins, one for each component of pollen-pistil recognition; in other words, self-incompatibility is a lock-and-key mechanism (Charlesworth et al. 2005). However, this does not alter the essential features of the use of self-incompatibility for hybrid production.

### ***Vegetables of Brassica oleracea***

Since the 1960s F<sub>1</sub> hybrid (single-cross) production has been the driving force behind the breeding of vegetable Brassicas: Brussels sprouts, cabbage, calabrese/broccoli and cauliflower. The aim has been to combine heterosis for high yield with high quality and uniformity in traits such as maturity that are associated with suitability for mechanical harvesting. Hybrids also provide built-in cultivar protection for commercial companies as they do not breed true on open-pollination. Initially hybrid Brassica breeding was based on use of the sporophytic self-incompatibility system. This was not always 100 % effective and sibs of the parents

were a problem in commercial crops. In other words, plants of a parent would successfully pollinate each other as well as plants of the other parent. A level of one or two per cent sibs was considered acceptable for cabbage, and higher levels in broccoli and cauliflower (Myers et al. 2012). Subsequently cytoplasmic male sterility (CMS) became the favoured method for hybrid production, once initial problems had been solved. Three systems involving the *B. oleracea* nuclear genome in other cytoplasms will be briefly presented.

The first system was developed by Pearson (1972) from the cross of *B. nigra* with broccoli, and then further developed in cabbage. However, it suffered from petaloidy and lack of development of the nectaries, and the male-sterile plants were therefore unattractive to pollinating bees. Nevertheless there is still some interest in improving this system (Kamiński et al. 2012).

The next system was developed from the sterility found in an unknown cultivar of Japanese radish (RR,  $2n = 2x = 18$ ) by Ogura (1968). Bannerot et al. (1974) crossed this 'Ogura' CMS radish (R cytoplasm) with cabbage (CC,  $2n = 2x = 18$ ) and in the fourth back-cross generation obtained totally male-sterile plants that were otherwise normal ( $2n = 18$ ). The flowers had only empty pollen grains or vestigial anthers. The original 'Ogura' cytoplasm from radish was unacceptable because it was associated with cold temperature chlorosis (Myers et al. 2012). This defect was corrected by replacing radish chloroplasts with those of *B. oleracea*. Walters et al. (1992) fused cauliflower (*B. oleracea* var. *botrytis*) protoplasts with Ogura male-sterile and *B. oleracea* male-fertile cytoplasms and produced plants with an array of organellar types. Those with Ogura mitochondria were male-sterile and those with *B. oleracea* chloroplasts were cold tolerant. Cauliflowers combining male-sterility and cold tolerance were found and propagated. Later, cold tolerant cytoplasmic male-sterile (CMS) cabbages (*B. oleracea* var. *capitata*) were produced 8 months after 66 independent fusions had been made of leaf protoplasts from fertile cabbage and cold-tolerant Ogura CMS broccoli lines (Sigareva and Earle 1997). The key features were selection of cabbage lines with a high frequency of plant regeneration from protoplasts; the use of gamma-irradiation (30 krad) to eliminate the nuclear DNA from the broccoli fusion partner; the use of flow cytometry and morphological traits to identify diploid cabbage plants; and the use of PCR primers for an Ogura CMS-specific mitochondrial DNA sequence. The use of irradiation and protoplast fusion is explained in more detail in Part IV of this book. Today 'Ogura' male-sterility is available in lines with good seed production and no apparent horticultural problems, but they are protected by patents taken out by various seed companies (Dixon 2007). Interestingly, Tanaka et al. (2012) have determined the complete nucleotide sequences of Ogura- and normal-type mitochondrial genomes in radish. The Ogura-type mitochondrial genome was highly rearranged compared with the normal-type one, with four unique regions, most of which were composed of known *Brassicaceae* mitochondrial sequences. This suggests that the regions unique to the Ogura-type genome were generated by integration and shuffling of pre-existing mitochondrial sequences during the evolution of *Brassicaceae*, and novel genes such as *orf138* could have been created in the process. It is the *orf138* protein that is involved in some way in producing the male-sterile phenotype.

A third source of male-sterile cytoplasm that has been transferred to *B. oleracea* using protoplast fusion (Cardi and Earle 1997) is the ‘Anand’ cytoplasm from *B. tournefortii*, which was first discovered by Rawat and Anand (1979) in *B. juncea*. Unlike ‘Ogura’, there were no initial problems with cold temperature chlorosis. Broccoli, cauliflower and cabbage lines with good seed yield have been produced (Dixon 2007). No fertility restoring genes for ‘Ogura’ or ‘Anand’ have been found to date in *B. oleracea* (or *B. rapa* or *B. napus*); but this does not matter in non flowering vegetable crops. All nuclear genotypes in normal cytoplasm are therefore maintainers, and the male-sterile version (female parent) of an inbred line is obtained simply by back-crossing the nuclear genome into the sterile cytoplasm. This can be done more quickly by substituting the nucleus of the inbred line into the sterile cytoplasm by protoplast fusion.

Breeding hybrids has also been accelerated by the production of doubled haploid parental lines using microspore culture, especially in the biennial crops such as cabbage (Myers et al. 2012).

## Onions and Carrots

### *Onions*

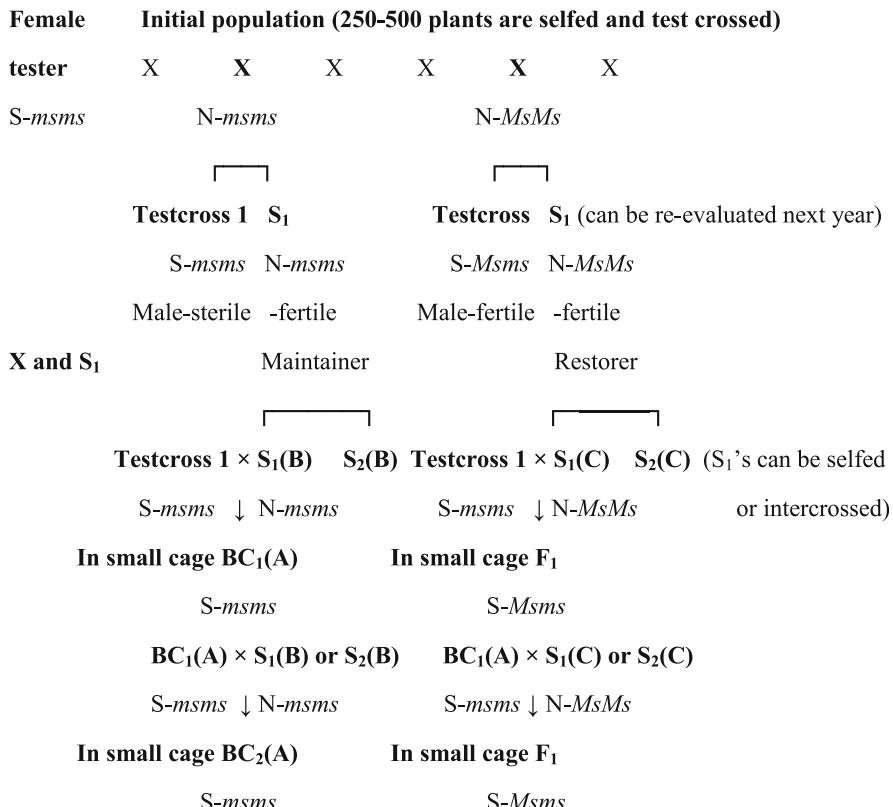
Dry onions (*Allium cepa*) rank second behind tomatoes in terms of worldwide importance as a vegetable crop, with 79 million tonnes produced in 2010 at latitudes from 5 to 60° in both hemispheres, of which 22 were produced in China and 15 in India. The bulbs of the common onion are normally single and plants reproduce from seeds or from seed-grown sets (small bulbs). This biennial, cross-fertilizing diploid species ( $2n = 2x = 16$ ) includes hundreds of traditional open-pollinated cultivars and modern F<sub>1</sub> hybrid ones, distinguished and categorized mainly by their bulb characteristics (Shigyo and Kik 2008). Cytoplasmic male sterility (CMS) was first discovered in 1925 in the cultivar Italian Red which was grown in the USA in California. Worldwide, it is still one of the most widely used sources of male sterility in F<sub>1</sub> hybrid production, and is designated CMS-S. The genetic basis of the system was worked out by Jones and Clarke (1943) and later research has been summarized by Shigyo and Kik (2008), including other sources of male sterility, one of which has been used in practice. The first commercial F<sub>1</sub> hybrids became available in the USA in the early 1950s and in the Netherlands’ export market in the late 1960s. Today the global temperate onion seed market comprises over 90 % F<sub>1</sub> hybrids, whereas the frequency of open-pollinated cultivars is still around 50–60 % in the one for the tropics (Shigyo and Kik 2008). The main advantages of F<sub>1</sub> hybrids over open-pollinated cultivars for seed companies are the in-built cultivar protection which forces farmers to buy fresh seed, and the ease of combining dominant traits. A detailed account of the goals and methods of hybrid onion breeding and seed production can be found in the review by Shigyo and Kik (2008). Here we will briefly consider some genetic aspects followed by practical aspects that differ from maize.

### ***The Use of CMS-S in Hybrid Onion Breeding***

CMS-S results from the combination of male-sterile cytoplasm (S) with the homozygous recessive genotype (*msms*) at a nuclear male-fertility-restoration locus (*Ms*). Male-sterile plants have composition S-*msms*, the thermostable sterility (no pollen under high temperatures) resulting from premature breakdown of the tapetum. Molecular markers are now available to detect CMS-S cytoplasm in germplasm collections in the presence of the *Ms* allele. Male fertile-plants with S cytoplasm have composition S-*Msm*s or S-*MsMs*, whereas male-fertile plants with normal cytoplasm (N) can have composition N-*Msm*s, N-*Msms* or N-*msms*. Male-sterile plants can be seed propagated by crossing S-*msms* (female) with N-*msms* (male) and harvesting seed from the female parent. They can then be used as the female parent in hybrid production. The male parent is the maintainer line whose nuclear genotype can be very similar to that of the female parent as a result of repeated backcrossing during maintenance. A high frequency of the *Ms* allele in germplasm of particular interest to a breeder makes finding a maintainer more difficult. As an onion crop does not flower, the unrelated male inbred parent used in hybrid production can have genotype N-*Msm*s or N-*msms*, i.e., it does not have to be a restorer as in maize. In fact it could have genotype S-*MsMs*. The effective use of CMS simply depends on consistent expression of the male-sterile and male-fertile phenotypes across environments and genetic backgrounds. However, as recently as 2010, Melgar and Havey (2010) have confirmed that the dominant *Ms* allele can show reduced penetrance, requiring that male-fertility restoration be scored over years to confidently assign genotypes at *Ms*.

### ***A Hybrid Onion Breeding Scheme***

The requirements for hybrid onion breeding are shown in Fig. 12.8. Details including photographs can be found in the review by Shigyo and Kik (2008). A source of CMS-S is needed. Then the plants in a population of onions (size 250–500) can be self-pollinated and also crossed to the source of CMS-S as female tester (S-*msms*). When a plant is selfed by bagging, the male-sterile female tester can be enclosed in the same bag, blowflies (or bees) being used for cross-pollination. The result will be pairs of S<sub>1</sub> and testcross families for evaluation. Selection of pairs is based on the performance of both the S<sub>1</sub> and testcross families. As both types of family are genetically variable, selection of good bulbs also takes place within the selected families. The inbreeding programme may only continue for two or three generations because of severe inbreeding depression. In other words, the parents of the hybrids are usually just partially inbred lines, and this is also the reason that doubled haploid production is not common practice. In Fig. 12.8 the essential features are shown for producing hybrids from S<sub>1</sub> families. The S<sub>1</sub> families can be divided into two groups on the basis of the male fertility of the testcross families, which are F<sub>1</sub>'s with S<sub>0</sub> plants. If all plants in a testcross family are male-sterile, the corresponding S<sub>1</sub> family (N-*msms*) can be used as a maintainer (B line). The testcross family (testcross 1) can be used as the female parent



**Fig. 12.8** F<sub>1</sub> hybrid onion breeding showing lines that survive to be female parent of hybrid (A), maintainer of female parent (B) and male parent of hybrid (C) which will be a restorer if *Ms* allele is frequent in initial population (as the backcrosses and selfs are genetically variable, selection is practised within them)

and the S<sub>1</sub> family as the male parent for production of the first backcross (BC<sub>1</sub>) generation in a small insect-proof cage, again using blowflies (or bees) for cross-pollination. This will be the female parent (A line) of the hybrid. With further backcrosses to the maintainer line (B) the female parent comes to resemble that line. Plants of the maintainer line can be intercrossed in a small cage to produce more seed, or bag-selfed to produce S<sub>2</sub> families for further selection before being used to produce the second backcross. The S<sub>1</sub> families that are not maintainers can be evaluated as the male parent (line C) of the hybrid by crossing with testcross 1 (S-msms). They can also be intercrossed in a small cage to produce more seed, or bag-selfed to produce S<sub>2</sub> families for further selection before being used as the male parent of the hybrid.

### ***Hybrid Onion Seed Production***

Shigyo and Kik (2008) also reviewed the issues that need to be considered in hybrid onion seed production. A few of the key ones are as follows. Production can be done

in a large glasshouse but is usually done in the field where honey bees are used for pollination. High quality seed with more than 95 % germination is important when onion seedlings are raised in cell trays before being planted. The bulb-to-seed method allows the grower to discard bulbs which are unacceptable in shape or size, whereas this is not possible in the faster seed-to-seed method that avoids bulb formation. It is important to maintain the purity of the female parent by carefully preventing any fertile male strains from getting mixed into the female plant lines. The seed production fields of different cultivars should be separated from each other by a linear distance of at least four kilometres to avoid the risk of unintended cross breeding. The male to female ratio in the seed production field should be within the range 1:2–1:4, depending on the reproductive capacity of the male parent.

## **Carrots**

Carrot (*Daucus carota*) is another globally important vegetable crop. It is an outcrossing, insect-pollinated diploid ( $2n=2x=18$ ) species with a biennial life cycle. The edible storage root forms and grows during the vegetative stage of its life cycle. Cold exposure (vernalization) is then required for the transition to flowering. The hermaphrodite flowers are usually protandrous and perfect, with each setting at most two seeds; but a single plant can produce hundreds of flowers on multiple umbels. Carrot breeding can be done on an annual cycle by raising a winter root crop in warmer production areas, harvesting and vernalizing roots for at least 6 weeks in refrigerated storage (1–5 °C), and then producing a summer seed crop in a cooler area (Simon et al. 2008). Small scale seed production can be done in insect-proof pollination cages into which flies or bees are released at flowering. For large scale seed production, cultivar integrity is ensured by geographical isolation of a few kilometres from other multiplications, and from wild carrots. Naturally occurring insect pollinators can be augmented with honeybee hives. Commercial seed production can be done seed-to-seed by sowing seed in the mid to late summer growing season so that plants are exposed to natural cold weather before roots form, and then flower the next year.

Following on from the success in onions, breeders sought a CMS system for carrot that could be used to produce  $F_1$ hybrid cultivars. The brown anthers form of male sterility was first discovered in the USA in cultivar Tendersweet and reported by Welch and Grimball (1947). It was subsequently found in several other cultivated and wild carrot sources. Anthers remain rudimentary, turn brown, and fail to produce mature pollen; but there have been some problems with stability of this form of male-sterility. Petaloid male sterility is the other form that has been found in North American wild carrot, first in 1953 (Simon et al. 2008). It is a dysfunction in flower development that results in the anthers being replaced by a second whorl of petals, presumably through modified mitochondrial genome organization and transcription affecting the expression of nuclear genes involved in flower formation. Male-fertility is restored by dominant alleles at either of two duplicate genes (Wolyn and Chahal 1998). Petaloid CMS is now the most widely used form for production of commercial

hybrids in North America. It is stable across a wide range of environments but can break down in late-season in some genetic backgrounds. Hybrids started to become available in the 1970s, and since the 1980s have been replacing open-pollinated cultivars worldwide. Today almost all cultivars bred and grown in North America, Europe, Korea and Japan are hybrids. The use of uniform hybrids suitable for mechanized production has reduced labour requirements and changed carrot growing from small scale market gardening to industrial production (Simon et al. 2008).

The CMS system in hybrid carrot breeding is similar to that for onions. The use of the CMS in testcrosses followed by assessment of male-fertility, divides the available germplasm into maintainers and restorers. Simon et al. (2008) report that it has been relatively easy to identify and select male sterility maintaining inbreds in carrot from a wide range of germplasm but the incidence of restorers varies widely. The male parent of the hybrid must produce adequate amounts of high quality pollen, whereas the female parent must be able to produce adequate amounts of seed with high germination rate. Simon et al. (2008) also review current breeding goals, methods and techniques; but we do not need to go into the details. While heterosis has significant positive effects upon many carrot production attributes, there are traits such as carotenoid and nutrient content where mid-parent values are observed in hybrids; and it is important to know this in designing a breeding programme. Furthermore, it is also important to remember that yield measurements in carrot are typical of all vegetable crops, with the focus on marketable and not total yield. Self-pollination is not restricted by incompatibility and carrot inbreds have typically been selfed for two to five generations. Although inbreeding depression can be severe, inbreds have been improved over the years and intercrossed to start new cycles of inbred development. Today uniform and vigorous hybrids can be produced by crossing fairly inbred and unrelated parents. Jagosz (2011) found that the marketable yield of 34 hybrids was 34 % higher than that of their 15 inbred parents, and that 13 hybrids displayed better-parent heterosis. Genetic distances between the inbred parents were estimated from AFLP markers, but the correlation ( $r = 0.34$ ) with better-parent heterosis for marketable yield was considered too low for predictive purposes.

## Rye

Rye (*Secale cereale*) is mainly a European cereal with about 75 % of global production in Russia, Belarus, Poland, Germany and Ukraine. Hybrid rye breeding started around 1970 in Germany (Geiger and Miedaner 2009). The production of inbred lines by self-pollination was made possible by the discovery of dominant self-fertility genes in various European germplasm. All presently listed hybrids are crosses between a cytoplasmic male sterile (CMS) single cross as seed parent and a restorer synthetic population as pollinator:

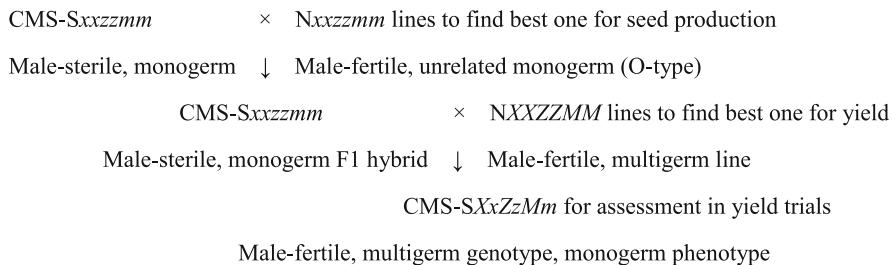
$$(A_{CMS} \times B) \times Syn_{RF}$$

The main source of CMS is known as P (Pampa) cytoplasm and is used because maintainers (B parent) were found at high frequency. In contrast, restorers with good pollen shedding proved harder to find but are now available (Geiger and Miedaner 2012). The parent lines A and B of the CMS single cross may be derived from the same or two different heterotic groups, but should be between unrelated homozygous lines in order to provide a uniform, vigorous and fertile female parent. The restorer synthetic is usually composed of two inbred lines to provide a longer pollen shedding period, but a single improved inbred line may be used in future. As in all hybrid breeding, parental line development and improvement is an ongoing process. Since rye is wind-pollinated careful isolation is required during commercial hybrid seed production to avoid genetic contamination from other sources. On large enough fields, the ratio of strips (6–8 m broad) of  $A_{CMS}$  to B can be 4:1, with seed being harvested from the  $A_{CMS}$  ones. The final step of seed production which provides certified seed for the farmer is done in a mixed stand of  $A_{CMS}B$  and  $Syn_{RF}$  in a ratio of about 95:5. This procedure reduces the cost of hybrid seed production compared with drilling alternate strips of the female and male parent, whilst the 5 %-admixture is too small to significantly affect the performance of the cultivar. The seed production field is surrounded by a 10-m-broad barrier of  $Syn_{RF}$  to protect it from mispollination, and samples of seed from all stages of hybrid production are authenticated by molecular marker fingerprinting. Today in Germany, hybrids cover about 70 % of the total rye acreage and the best ones out yield the best open-pollinated cultivars by 15–20 %. Nevertheless, there is still a demand for open-pollinated cultivars for some farming systems and markets.

## Sugar Beet

Sugar beet (*Beta vulgaris*) is an outbreeding crop with self-fertilization normally prevented by a gametophytic self-incompatibility system. We met the crop back in Chap. 1 under selection for altered chemical composition. Hybrid sugar beet (and fodder beet) needs to be seen in the context of two other major developments in the breeding and growing of the crop, as explained in the review by Biancardi et al. (2010). Tetraploid sugar beets ( $2n = 4x = 36$ ) were produced and evaluated following the discovery that the doubling of chromosome number could be induced by treatment with colchicine (Blakeslee and Avery 1937). Diploid ( $2n = 2x = 18$ ) and tetraploid sugar beet were easily crossed to produce triploid hybrids which displayed better sugar yields than their parental averages. Monogerm cultivars became possible following the discovery in 1948 of plants with single flowers that produced monogerm seeds, a trait that was inherited as a simple Mendelian recessive  $mm$  (Savitsky 1950). This eliminated either the manual thinning of seedlings from multigerm seed or the mechanical processing of multigerm glomerules into single seeds. The first monogerm lines were also self-fertile as a result of a dominant gene  $S_f$  (Savitsky 1950); a trait that proved useful for the development of inbred lines.

Cytoplasmic-genetic male sterility was discovered by Owen (1945). The occurrence of male-sterile cytoplasm (S) required the presence of two homozygous



**Fig. 12.9** Breeding scheme for diploid monogerm three-way cross hybrid ( $A \times B$ )  $\times C$  of sugar beet, where Sxxzz is required for male-sterility and mm in the female parent for monogerm seed

recessive nuclear genes *xxzz*: CMS-Sxxzz. The reproduction of CMS lines required the employment of maintainers with normal (N) cytoplasm and nuclear genes *xxzz*. The monogerm trait was transferred by six to eight generations of backcrossing to the maintainers which were now called O-types. The employment of CMS monogerm lines resulted in new types of hybrid cultivars, as explained by Biancardi et al. (2010). As the harvested crop comes from vegetative and not sexual reproduction, only the female parent has to be monogerm. Pollination by a multigerm male line gives seed that is genetically multigerm, but phenotypically monogerm. For some decades it was triploid hybrids that had most commercial success, at least in Europe, as a result of their superior sugar yields. Diploid CMS  $F_1$  female parents were pollinated by tetraploid multigerm lines or hybrids ( $2x$  CMS *mm*  $F_1 \times 4x$  N *MM*) to give triploid three-way cross or double cross hybrids ( $3x$  CMS *Mm*). The diploid  $F_1$  parents were produced by crossing CMS lines with different O-type lines. Hybrids were also made with pollinators reproduced by intercrossing (families) and these were therefore topcross hybrids. During the last 25 years improved diploid pollinators have been developed from a broad genetic base so that diploid hybrids are now becoming prevalent in Europe and elsewhere (Biancardi et al. 2010). Most are three-way cross ( $2x$  CMS *mm*  $F_1 \times 2x$  N *MM*) rather than single cross hybrids because of the lower seed production of CMS inbred lines. A breeding scheme is outlined in Fig. 12.9.

Details of commercial hybrid seed production can be found in the review by Biancardi et al. (2010). A few key points are as follows. A row of pollinators is sown or planted to every three or four rows of the CMS line and seed harvested from the latter, the pollinator having been removed after flowering. This allows inspection before flowering to eliminate any fertile, anomalous or off-type plants. Seed can be sown directly at the site of production in autumn, or in a nursery from where small roots (stecklings) are transplanted into the seed production field in spring, sometimes after storage and vernalization in a cold room over winter. Topping the seed-bearing stalks favours the development of lateral branches, improves the uniformity of seed size and is useful for synchronizing flowering of pollinators and seed bearers. Isolation from other pollen sources is important because sugar beet is wind pollinated. Unwanted pollen can come from other seed crops as well as from wild, bolted and weed beets; the latter have been shown to receive pollen released up to 9.6 km away.

# **Chapter 13**

## **Inbred Line Cultivars and Mixtures from Hybridization and Inbreeding**

### **Introduction**

In this chapter we are going to look at strategic questions in the breeding of inbred line cultivars of crops whose natural sexual reproduction involves a high level of self-pollination. In other words, they are tolerant of inbreeding and suffer little inbreeding depression. We are going to concentrate on the use of artificial hybridization to generate genetic variation from which superior inbred line cultivars can be produced. We will look at the choice of parents, the number and choice of crosses to be made, the size of progenies and the theory of selection within progenies. As in previous chapters, selection within progenies is multi-trait and multistage, but we now have the additional complication of inbreeding. More detailed information on particular crops can be found in the *Handbook of Plant Breeding* (Prohens et al. 2008). We will finish the chapter with a consideration of cultivar mixtures.

### **Pure-Line Selection**

Analyses of the pedigrees of cultivars of naturally inbreeding crops from the late nineteenth century through to the present day reveals a transition in breeding approach, from selecting true breeding lines from genetically variable landraces to the use of deliberate artificial (hand) hybridization in order to make further progress. Two examples will suffice. The first comes from the choice of cultivars made by Russell et al. (2000) in their retrospective analysis of spring barley germplasm development from “foundation genotypes” to currently successful cultivars. The seven cultivars released from 1884 to 1920 were all selections, five from landraces and two from cultivars. The 20 from 1921 to 1960 comprised seven selections, again five from landraces and two from cultivars, one mutation, and 16 from crosses. Finally 50 out of the 51 cultivars released from 1961 to 1990 were

from crosses, with just one in 1972 which was a selection for disease resistance. My second example is from traditional Basmati rice where pure-line breeding was adopted in the twentieth century to improve locally adapted land races by selecting and multiplying desirable plants identified in natural populations (Rao et al. 2012). Basmati 370 was a pure-line selection of Dehradooni Basmati, a locally adapted land race, identified at Kala Shah Kaku research station in Pakistan. It was therefore recognized as Traditional Basmati, became the bench mark for quality, and brought an export boom to both India and Pakistan. Further progress was made by cross-breeding Basmati cultivars with other types and selecting for Basmati quality, but the resulting cultivars were classed as Evolved rather than Traditional Basmati. For example, Haryana Basmati-1 is a semi-dwarf aromatic rice cultivar from the cross between Sona and Basmati 370 (Panwar et al. 1991).

### ***Artificial Hybridization***

During the twentieth century, cross-pollination followed by pedigree inbreeding with selection, single seed descent, the production of doubled haploids, or bulk population selection, became established as very successful breeding methods for naturally inbreeding crops. Parents are chosen which complement one another for desirable characteristics. The cross-pollination creates the potential genetic variation from which inbred line cultivars superior to their parents can be produced by selection during one or more stages of the breeding process. Later in this chapter we will examine the effectiveness of selection in these breeding methods. First, however, we need to consider the overall size of the programme required in terms of the choice of parents, the number and choice of crosses to be made, and the size of progenies. The decisions are straightforward if we know the sources and number of alleles that need to be combined in a new cultivar to achieve our objectives. The selection process is also straightforward if the alleles can be uniquely recognized at the DNA level or through their expression. For example, if we need to bring together four alleles from four different sources, this cannot be done in a single round of pair crosses; rather two rounds of crossing are required. If four desired alleles at four unlinked loci are segregating in the  $F_2$  of a cross, one 16th of the recombinant inbred lines that can be produced will be the desired true-breeding line. In other words, a small progeny size is adequate. With seven unlinked loci, the frequency is 1 in 128, and the reader will recall that Mendel produced all combinations of his seven single factor traits without too much difficulty. However, if 20 unlinked loci are segregating, the frequency of the desired line will be 1 in  $2^{20}$ , or 1 in 1,048,576, and a breeder would anticipate more than one cycle of crossing and selection. Likewise, if desirable and undesirable alleles were tightly linked in coupling ( $Ab/aB$ ), more than one generation might be required for the necessary genetic crossing over to occur. For example, Witcombe and Virk (2001) presented data from the  $F_2$  of a pearl millet cross showing that linkage groups are often inherited unchanged from each parent and that it is the norm for large regions of the

genome to be inherited unchanged. In the coming years the precision of plant breeding will continue to increase as more and more economically important genes are mapped and desirable alleles are recognized. In the meantime, however, we are still going to need a crossing strategy for quantitative traits where the number of desired alleles is unknown and individual alleles are not recognized. Although the  $F_1$  hybrids between pairs of parents could be inter-crossed to bring together desirable alleles from four parents, we will concentrate on the methods of producing inbred line cultivars from the  $F_1$  hybrids.

## Genetic Variation Between and Within Crosses

### Theory

It is useful to start with some theory. Let us assume that we have a collection of homozygous lines that differ at two unlinked loci, with either genotype  $AA$  or  $aa$  at the first locus and either genotype  $BB$  or  $bb$  at the second locus. Furthermore, let us assume that the frequencies of these genotypes are  $p$  and  $q$  and  $u$  and  $v$ , and that the two loci are in equilibrium so that the frequencies of the four genotypes  $AABB$ ,  $AAbb$ ,  $aaBB$  and  $aabb$  are  $pu$ ,  $pv$ ,  $qu$  and  $qv$ , respectively. If the homozygous lines are crossed in all possible combinations (a diallel set of crosses) and all possible recombinant inbred lines are produced from the resulting  $F_1$ 's, the variation among the  $F_\infty$  lines can be partitioned into between and within crosses as shown in Table 13.1. The genetic values of the four genotypes are:

$AABB$	$AAbb$	$aaBB$	$aabb$
$a_A + a_B + aa_{AB}$	$a_A - a_B - aa_{AB}$	$-a_A + a_B - aa_{AB}$	$-a_A - a_B + aa_{AB}$

where  $a_A$  and  $a_B$  are the additive effects at the two loci and  $aa_{AB}$  (this is not 'a' multiplied by 'a') is the interaction between them (complementary epistasis:  $a_A = a_B = aa_{AB}$ ; duplicate epistasis:  $a_A = a_B = -aa_{AB}$ ).

The means ( $\mu$ ), variances and covariances are calculated in the usual way for frequencies ( $f$ ):

$$\text{Variance of } x's = \sum (x - \mu_x)^2 / n = \sum f x^2 - (\mu_x)^2 \text{ where } \Sigma f = 1 \text{ and } \mu_x = \Sigma f x$$

$$\text{Covariance of } x's \text{ and } y's = \Sigma (x - \mu_x)(y - \mu_y) / n = \Sigma f x y - (\mu_x)(\mu_y) \text{ where } \Sigma f = 1, \mu_x = \Sigma f x \text{ and } \mu_y = \Sigma f y.$$

The mean of the parents equals the mean of all of the  $F_\infty$  lines:  $a_A(p-q) + a_B(u-v) + aa_{AB}(p-q)(u-v)$

**Table 13.1** Variation between and within crosses for two unlinked loci ( $f$  is frequency)

Parent	Parent	Difference	Mid-par	$f$	$F_{\infty}$	$F_{\infty}$ mean	$F_{\infty}$ var
AABB	AABB	0	$a_A + a_B + aa_{AB}$	$p^2 u^2$	AABB	$a_A + a_B + aa_{AB}$	0
AABB	AAAb	$2(a_B + aa_{AB})$	$a_A$	$2p^2 uv$	$\frac{1}{2}AABB$	$(a_B + aa_{AB})^2$	$(a_B + aa_{AB})^2$
AAAb	AAAb	0	$a_A - a_B - aa_{AB}$	$p^2 v^2$	$\frac{1}{2}AAAb$	$a_A$	$a_A$
AABB	aaBB	$2(a_A + aa_{AB})$	$a_B$	$2pq u^2$	$AAAb$	$a_A - a_B - aa_{AB}$	0
AABB	aaBB	$2(a_A + a_B)$	$aa_{AB}$	$2pq u v$	$\frac{1}{2}AAAB$	$a_B$	$(a_A + aa_{AB})^2$
AAAb	aaBB	$2(a_A - a_B)$		$2pq u v$	$\frac{1}{2}aaBB$	0	$(a_A + aa_{AB})^2$
AAAb	aaBB	$2(a_A - a_B)$		$2pq u v$	$\frac{1}{2}AAAB$	$a_A^2 + a_B^2 + aa_{AB}^2$	$a_A^2 + a_B^2 + aa_{AB}^2$
AAAb	aaBB	$2(a_A - a_B)$		$2pq u v$	$\frac{1}{2}AAAb$	$a_A^2 + a_B^2 + aa_{AB}^2$	$a_A^2 + a_B^2 + aa_{AB}^2$
AAAb	aaBB	$2(a_A - a_B)$		$2pq u v$	$\frac{1}{2}AAAb$	0	$a_A^2 + a_B^2 + aa_{AB}^2$
AAAb	aaBB	$2(a_A - a_B)$		$2pq u v$	$\frac{1}{2}AAAb$	$a_A^2 + a_B^2 + aa_{AB}^2$	$a_A^2 + a_B^2 + aa_{AB}^2$
AAAb	aaBB	$2(a_A - a_B)$		$2pq u v$	$\frac{1}{2}AAAb$	$-a_B$	$(a_A - aa_{AB})^2$
aaBB	aaBB	0	$-a_A + a_B - aa_{AB}$	$q^2 u^2$	$\frac{1}{2}aaBB$	$-a_A + a_B - aa_{AB}$	0
aaBB	aaBB	$2(a_B - aa_{AB})$	$-a_A$	$2q^2 uv$	$\frac{1}{2}aaBB$	$-a_A$	$(a_B - aa_{AB})^2$
aaBB	aaBB	0	$-a_A - a_B + aa_{AB}$	$q^2 v^2$	$\frac{1}{2}aaBB$	$-a_A - a_B + aa_{AB}$	0

Total variance of $F_\infty$ lines:	$4pq[a_A + (u - v)aa_{AB}]^2 + 4uv[a_B + (p - q)aa_{AB}]^2 + 16pquvaa_{AB}^2$
Between crosses:	$2pq[a_A + (u - v)aa_{AB}]^2 + 2uv[a_B + (p - q)aa_{AB}]^2 + 4pquvaa_{AB}^2$
Average within crosses:	$2pq[a_A + (u - v)aa_{AB}]^2 + 2uv[a_B + (p - q)aa_{AB}]^2 + 12pquvaa_{AB}^2$
Variance of mid-parent values:	$2pq[a_A + (u - v)aa_{AB}]^2 + 2uv[a_B + (p - q)aa_{AB}]^2 + 8pquvaa_{AB}^2$

Covariance of mid-parent values and  $F_\infty$  means of crosses equals the between crosses variance. Hence the regression ( $b$ ) of the  $F_\infty$  means of crosses on the mid-parent values is:

$$\frac{\left\{ 2pq[a_A + (u - v)aa_{AB}]^2 + 2uv[a_B + (p - q)aa_{AB}]^2 + 4pquvaa_{AB}^2 \right\}}{\left\{ 2pq[a_A + (u - v)aa_{AB}]^2 + 2uv[a_B + (p - q)aa_{AB}]^2 + 8pquvaa_{AB}^2 \right\}} \text{ divided by}$$

### Conclusions from Theory

The main conclusions from this theory which require discussion are as follows.

1. The variation between and within crosses is similar, with the average within variance greater than the between in the presence of epistasis but identical in its absence. Therefore in selecting the best cultivar, roughly equal contributions are expected from selecting the best cross and then the best line within that cross. However, there is variation in the within cross variance so that the best cultivar is expected to come from the cross with the highest variance as well as the highest mean, so that the within cross variance makes an additional contribution. However, we shall see that it is difficult to detect real differences in the within cross variances because they have large standard errors. Furthermore, as the number of loci increases, for unrelated parents from populations in equilibrium, extreme variances become rare. If there are  $n$  loci, the chances of  $AABBCC\dots NN \times AABBCC\dots NN$  with zero variance and  $AABBCC\dots NN \times aabbcc\dots nn$  with very high variance are both very low. Much more likely are crosses with parents differing at around half their loci, such as  $AABBCC\dots NN \times aaBBcc\dots NN$ , so that the variation observed in the variances is relatively small. We shall see later in the book that when highly adapted parents are crossed with very unadapted ones carrying desirable traits, a backcrossing programme is more likely to ensue.
2. The mid-parent value is a perfect predictor of the cross mean in the absence of epistasis and is still a good predictor in the presence of epistasis.
3. In contrast, there is not a good predictor of the within cross variance. For 8 out of the 10 crosses in the table, half the difference between the parents does predict the square root of the variance, but these are the crosses where the parents are either identical or just differ at one locus. We need to consider parents that differ at more than one locus. The variance for  $AABB \times aabb$  is the same as for  $AAbB \times aaBB$  ( $a_A^2 + a_B^2 + aa_{AB}^2$ ), but half the differences are ( $a_A + a_B$ ) and

( $a_A - a_B$ ), respectively, so that the former is larger than the latter which will always be smaller than the square root of the variance. Recombinant inbred lines superior to their parents (transgressive segregation) come from a dispersion of desirable alleles between the parents ( $AAbb$  and  $aaBB$ , but usually over many more loci), but we cannot detect such pairs of parents from their phenotypes.

## ***Experimental Results***

These conclusions from theory were tested by Utz et al. (2001) with 22 lines randomly derived from each of 30 winter wheat crosses produced by a factorial mating of five high yielding with six high baking quality cultivars. This is perhaps more realistic than crossing the 11 cultivars in all possible combinations. The parents and lines were evaluated in four environments for seven agronomic and quality traits. The mid-parent values were good predictors of cross means despite small deviations due to epistatic effects for most traits. The correlation ( $r$ ) between cross means and mid-parent values was greater than 0.70 for six of the seven traits. For the one trait with a lower correlation, the variation among cross means was not statistically significant. The genetic variance between the means of crosses was mostly smaller than the average segregation variance among lines within crosses. However, a significant difference between these variances could be confirmed only for one trait because of the large standard errors associated with the variance component estimates. Hence it is difficult to test the hypothesis that a smaller genetic variance between mid-parent values might be expected when all crosses are between adapted and unadapted parents, as here, or when all parents are phenotypically similar because they are highly adapted to the same target environments and end uses. The range of estimates of the segregation variance among lines within crosses was high for all traits, but the heterogeneity was statistically significant for only two out of the seven traits. Furthermore, no significant associations were found between these within cross variances and any measure of differences between the parents, including the squared phenotypic difference between parents. For reasons explained earlier, there are theoretical reasons for poor correlations between the true values of parental differences and within cross variances, as well as problems with the high standard errors of the latter. In future, the correlations may be improved by the application of molecular markers to estimating the relationship between parents. However, the molecular markers will need to be closely linked with the QTL involved in trait expression.

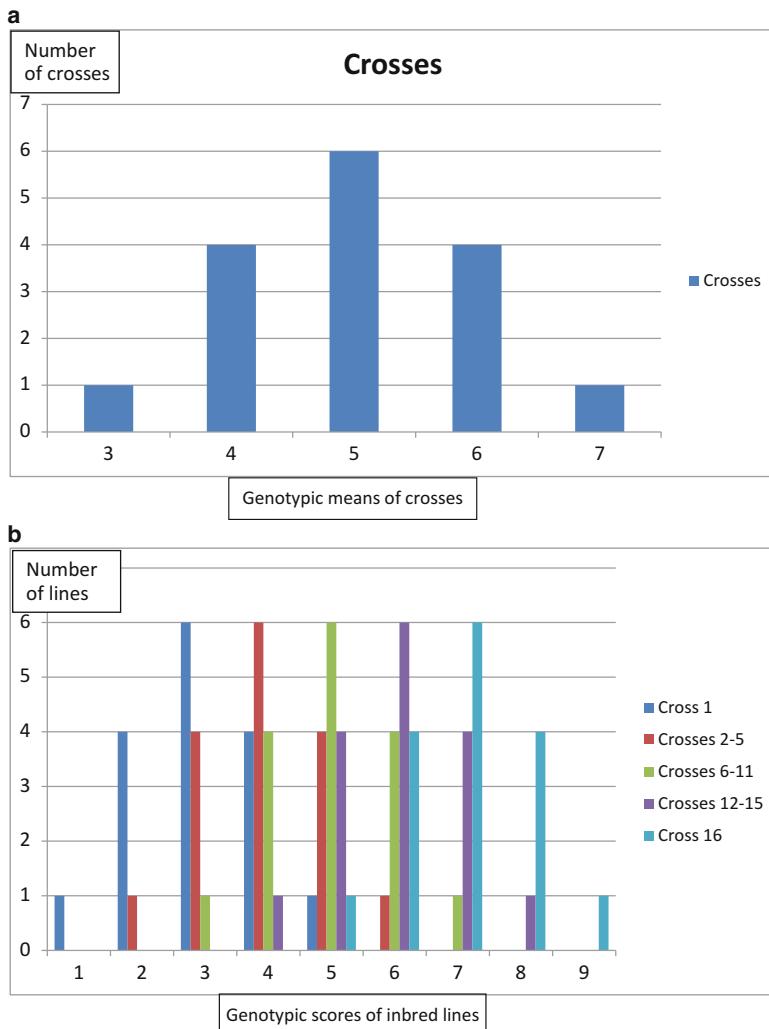
In a companion study, Bohn et al. (1999) genotyped the parental lines with 21 SSR (simple sequence repeat) and 16 AFLP (amplified fragment length polymorphism) primer-enzyme combinations but no significant association was found between genetic distances on the basis of these markers and within cross variance for all seven traits. Utz et al. (2001) concluded that prediction of within cross variances remained an unsolved problem. They also looked at variation in the usefulness of the crosses, where the usefulness of a cross was defined as the cross

mean (mean of all possible homozygous lines from cross) plus the expected selection gain (heritability  $\times$  selection differential) when a fraction of the homozygous lines was selected with regard to a given selection criterion. The variation in usefulness is thus determined by variation in the cross means and variation in the square root of within cross variances, assuming no variation in the intensity of selection times the square root of the heritability. Utz et al. (2001) found that for most traits, the variation in cross means was large enough for the additional contribution of the square root of the within cross variance to be negligible. This can be understood from a very simple example. If we have equal frequencies of the inbred lines  $AABB$ ,  $AAbb$ ,  $aABb$  and  $aabb$ , with genetic values  $2a$ ,  $0$ ,  $0$ , and  $-2a$ , and cross them in all possible combinations, the variance of cross means will be  $a^2$  and the variance of the within cross standard deviations will be  $0.272a^2$  (from Table 13.1). But if there were genetic differences at eight rather than two loci, the respective values would be  $4a^2$  and  $0.149a^2$ . In other words, the ratio of these two sources of variation rapidly increases as the number of loci increase.

More recently Hung et al. (2012b) tested the reliability of genotypic and phenotypic distance estimators between pairs of maize inbred lines to predict genotypic variation for quantitative traits within families derived from biparental crosses. They developed 25 families each composed of ~200 random recombinant inbred lines from crosses between a common parent, inbred B73, and 25 diverse maize inbreds. Parents and families were evaluated for 19 quantitative traits in up to 11 environments. Genetic distances (GDs) among parents were estimated with 44 simple sequence repeat and 2303 single-nucleotide polymorphism markers. GDs among parents had no predictive value for progeny variation. In contrast, there was a positive correlation ( $r > 0.7$ ) between phenotypic parental distances and within-family genetic variance estimates for five traits: days to anthesis, days to silk, upper leaf angle, tassel length and tassel primary branches. Overall, prediction of within cross variances remained elusive.

## Initial Choice of Crosses

The same arguments can be used here for inbred line cultivars as were deployed in Chap. 11 for vegetatively propagated cultivars. As one can predict cross means, but not variances, the initial choice of crosses needs to be based on differences in mid-parent values and the assumption that all crosses have the same within cross variance. Hence it is very important to have as much information as possible about potential parents. Simmonds (1996) demonstrated that the proportion of segregates within crosses superior to standards falls away quickly with increasing rank (from best to worst) of crosses, where realistic standards have means close to the means of the best crosses. Hence the best few crosses contribute most of the superior segregates and there is no need to make and evaluate other crosses from parents with clearly inferior mid-parent values. These points can be illustrated with the theoretical example shown in Fig. 13.1, which is based on the binomial distribution



**Fig. 13.1** Variation between and within 16 crosses where 16 inbred lines have been produced from each cross. (a) Histogram of 16 crosses showing variation between their means. (b) Histogram of 16 inbred lines from each cross with a different mean

for simplicity. Sixteen inbred lines are produced from each of 16 crosses. The means of the 16 inbred lines are 3 for the first cross, 4 for the next four, 5 for the next six, 6 for the next four and 7 for the final cross. The variance of the cross means is 1.0. The variation between the inbred lines within a cross is the same for each cross, and also has a variance of 1.0. Hence the variation between the crosses is the same as the variation within crosses. If we decide to select inbred lines with scores greater than the mean of the best cross (i.e.  $>7.0$ ), then this best cross will contribute five inbred lines, the four crosses ranked second equal will each contribute one line, and

the other 11 crosses will not contribute at all. Simmonds (1996) also provided evidence from five data sets (barley, two for wheat, oats and rice) of normality of the distribution of cross means, an assumption in his considerations.

In an SSD breeding programme for swedes, Bradshaw et al. (2009a) found that 15 crosses made an unequal contribution to the 50 families selected at  $F_6$ , with most in the range of 1–5, but with 15 and 11 from crosses 2 and 3, respectively, and none from crosses 12 and 14. The superiority of cross 2 over cross 12 was clear from their histograms of dry-matter yield, and could have been predicted from the mid-parent values of 9.63 t/ha (rank fourth) and 7.07 t/ha (rank 13th), respectively, compared with their  $F_6$  means of 10.43 t/ha and 7.26 t/ha, respectively. In fact, if the crosses ranked 1–7 on mid-parent yields had been selected, they would have included six out of the top seven crosses at  $F_6$ .

## Discarding Crosses at $F_1$ and $F_3$

Once crosses have been made, if differences in their means and variances can be detected, then it may be possible to safely discard whole crosses and concentrate the search for a new cultivar on the remaining more promising ones. Based on a quantitative genetics analysis of modern Indica cultivars of rice in Sri Lanka, Lawrence and Senadhira (1998) advocated a breeding procedure in which selection was practised between crosses at  $F_1$  and  $F_3$ , and within crosses at  $F_6$ . They argued that a trial of parents and their  $F_1$ 's would allow the detection of heterotic crosses, and in the absence of overdominance, this would imply a dispersion of some desirable genes between the parents that could be combined in their offspring. Hence the most promising crosses could be identified and retained, namely those combining a high mid-parental value with better parent heterosis (i.e. the  $F_1$  higher yielding than the better parent). At  $F_3$  it is possible to make predictions about the proportion of recombinant inbred lines whose means will exceed a desired target for individual traits and combinations of traits. Hence once again there should be the opportunity to identify and retain the most promising crosses from which to seek new cultivars at  $F_6$ . Furthermore, the predictions should provide guidance on the sizes of  $F_6$  populations required to achieve targets for improvements. The predictions are obtained from the standardized normal distribution, based on estimates of the mean ( $m$ ) and standard deviation ( $SD$ ) of the recombinant inbred lines ( $F_\infty$  generation), so that the target mean minus  $m$ , divided by  $SD$ , can be used as a normal probability integral (Kearsey and Pooni 1996). However, as the predictions are only approximate, we need to decide how much confidence can be placed in them, and for this purpose it is useful to look at the theory for two unlinked loci with epistasis (Box 13.1) and two linked loci without epistasis (Box 13.2). We will not consider the more complicated situation of two linked loci with epistasis.

**Box 13.1: F<sub>3</sub> family means: two unlinked loci with epistasis**

F <sub>2</sub> genotype	Genetic value	Frequency × 16	F <sub>3</sub> family mean
AABB	$a_A + a_B + aa_{AB}$	1	$a_A + a_B + aa_{AB}$
AABb	$a_A + d_B + ad_{AB}$	2	$a_A + \frac{1}{2}(d_B + ad_{AB})$
AAAb	$a_A - a_B - aa_{AB}$	1	$a_A - a_B - aa_{AB}$
AaBB	$a_B + d_A + ad_{BA}$	2	$a_B + \frac{1}{2}(d_A + ad_{BA})$
AaBb	$d_A + d_B + dd_{AB}$	4	$\frac{1}{2}(d_A + d_B) + \frac{1}{4}dd_{AB}$
Aabb	$-a_B + d_A - ad_{BA}$	2	$-a_B + \frac{1}{2}(d_A - ad_{BA})$
aaBB	$-a_A + a_B - aa_{AB}$	1	$-a_A + a_B - aa_{AB}$
aaBb	$-a_A + d_B - ad_{AB}$	2	$-a_A + \frac{1}{2}(d_B - ad_{AB})$
aabb	$-a_A - a_B + aa_{AB}$	1	$-a_A - a_B + aa_{AB}$

The means ( $\mu$ ) and variances are calculated in the usual way for frequencies ( $f$ ):

$$\text{Variance of } x's = \sum (x - \mu_x)^2 / n = \sum fx^2 - (\mu_x)^2 \text{ where } \sum f = 1 \text{ and} \\ \mu_x = \sum fx$$

The mean of the F<sub>3</sub> generation is:

$$(d_A + d_B)/4 + dd_{AB}/16$$

(summed over loci and pairs of loci =  $m + [d]/4 + [dd]/16$ )

The variance of F<sub>3</sub> family means is:

$$\left[ (a_A + \frac{1}{4}ad_{AB})^2 + (a_B + \frac{1}{4}ad_{BA})^2 \right] / 2 + \left[ (d_A + \frac{1}{4}dd_{AB})^2 + (d_B + \frac{1}{4}dd_{AB})^2 \right] / 16 \\ + [aa_{AB}^2]/4 + [ad_{AB}^2 + ad_{BA}^2]/32 + [dd_{AB}^2]/256$$

(summed over loci and pairs of loci =  $V_A^* + V_D^*/4 + V_{AA}^* + V_{AD}^*/4 + V_{DD}^*/16$ , where

$$V_A^* = \sum \left[ (a_A + wad_{AB})^2 + (a_B + wad_{BA})^2 \right] / 2$$

$$V_D^* = \sum \left[ (d_A + wdd_{AB})^2 + (d_B + wdd_{AB})^2 \right] / 4$$

$$V_{AA}^* = \sum [aa_{AB}^2] / 4$$

$$V_{AD}^* = \sum [ad_{AB}^2 + ad_{BA}^2] / 8$$

$$V_{DD}^* = \sum [dd_{AB}^2] / 16$$

and ‘w’ =  $\frac{1}{4}$  for the variance of F<sub>3</sub> family means).

(continued)

**Box 13.1** (continued)

(for completeness, the  $F_2$  variance is

$$V_A^* + V_D^* + V_{AA}^* + V_{AD}^* + V_{DD}^*, \text{ with } w = \frac{1}{2}$$

### **Use of $F_3$ Generation to Predict Recombinant Inbred Lines at $F_\infty$ Generation**

In summary, the mean of the  $F_3$  generation for two unlinked genes with epistasis is:  $m + (d_A + d_B)/4 + dd_{AB}/16$  compared with  $m$  for the  $F_\infty$  generation, and twice the variance of the  $F_3$  generation is

$$\left[ (a_A + \frac{1}{4}ad_{AB})^2 + (a_B + \frac{1}{4}ad_{BA})^2 \right] + \left[ (d_A + \frac{1}{4}dd_{AB})^2 + (d_B + \frac{1}{4}dd_{AB})^2 \right]/8 + [aa_{AB}^2]/2 + [ad_{AB}^2 + ad_{BA}^2]/16 + [dd_{AB}^2]/128 \text{ compared with } [a_A^2 + a_B^2] + aa_{AB}^2 \text{ for the } F_\infty \text{ generation,}$$

where  $a_A$  and  $a_B$  are the additive effects at the two loci and  $aa_{AB}$  is the interaction between them,  $d_A$  and  $d_B$  are the dominance effects at the two loci and  $dd_{AB}$  is the interaction between them, and  $ad_{AB}$  and  $ad_{BA}$  are the interactions between additive and dominance effects (classical complementary epistasis when dominance is for the higher score:  $a_A = a_B = d_A = d_B = aa_{AB} = ad_{AB} = ad_{BA} = dd_{AB}$ ; duplicate epistasis:  $a_A = a_B = d_A = d_B = -aa_{AB} = -ad_{AB} = -ad_{BA} = -dd_{AB}$ ; complementary epistasis when dominance is for the lower score:  $a_A = a_B = -d_A = -d_B = -aa_{AB} = ad_{AB} = ad_{BA} = -dd_{AB}$ ; duplicate epistasis:  $a_A = a_B = -d_A = -d_B = aa_{AB} = -ad_{AB} = -ad_{BA} = dd_{AB}$ ; in other words,  $d$  and  $dd$  have the same sign with complementary epistasis and opposite signs with duplicate epistasis).

Hence in the presence of dominance, the  $F_3$  mean as an estimate of  $m$  will be biased by  $(d_A + d_B)/4$  and twice the  $F_3$  variance as an estimate of  $[a_A^2 + a_B^2]$  will be inflated by  $[d_A^2 + d_B^2]/8$ . In the presence of epistasis the  $F_3$  mean will also be biased by  $dd_{AB}/16$  and the variance will be over or under estimated, depending on the type of epistasis. Thus for classical complementary epistasis with dominance for the higher score we have, in terms of  $a_A^2$ , a variance of 531/128 compared with 384/128 at  $F_\infty$  (over estimate), and for classical duplicate epistasis a variance of 243/128 (under estimate). The extent of bias depends on the extent of epistasis in the particular cross. Kearsey and Pooni (1996) reported that experimental studies have shown that the impact of epistasis is greater on the means than the variances, and that the variance obtained from the  $F_3$  generation (or triple-test-cross) can be used to calculate a satisfactory estimate of  $2V_A^* + 4V_{AA}^*$  (the  $F_\infty$  variance).

**Box 13.2: Two linked loci without epistasis**

Let us start with two linked loci  $A$  and  $B$ , with recombination frequency  $r$ . We need to consider the two possible  $F_1$ 's,  $AB/ab$  and  $Ab/aB$ , with the increasing

(continued)

**Box 13.2** (continued)

alleles in coupling and repulsion, respectively. The frequencies of the gametes from  $AB/ab$  are  $AB \frac{1}{2}(1-r)$ ,  $ab \frac{1}{2}(1-r)$ ,  $Ab \frac{1}{2}r$  and  $aB \frac{1}{2}r$ , whereas from  $Ab/aB$  they are  $Ab \frac{1}{2}(1-r)$ ,  $aB \frac{1}{2}(1-r)$ ,  $AB \frac{1}{2}r$  and  $ab \frac{1}{2}r$ .

$F_1 AB/ab$			$F_1 Ab/aB$		
$F_2$	Frequency	Genetic value	$F_2$	Frequency	Genetic value
$AABB$	$\frac{1}{4}(1-r)^2$	$a_A + a_B$	$AABB$	$\frac{1}{4}r^2$	$a_A + a_B$
$AABb$	$\frac{1}{2}r(1-r)$	$a_A + d_B$	$AABb$	$\frac{1}{2}r(1-r)$	$a_A + d_B$
$AAbb$	$\frac{1}{4}r^2$	$a_A - a_B$	$AAbb$	$\frac{1}{4}(1-r)^2$	$a_A - a_B$
$AaBB$	$\frac{1}{2}r(1-r)$	$a_B + d_A$	$AaBB$	$\frac{1}{2}r(1-r)$	$a_B + d_A$
$AB/ab$	$\frac{1}{2}(1-r)^2$	$d_A + d_B$	$AB/ab$	$\frac{1}{2}r^2$	$d_A + d_B$
$Ab/aB$	$\frac{1}{2}r^2$	$d_A + d_B$	$Ab/aB$	$\frac{1}{2}(1-r)^2$	$d_A + d_B$
$Aabb$	$\frac{1}{2}r(1-r)$	$-a_B + d_A$	$Aabb$	$\frac{1}{2}r(1-r)$	$-a_B + d_A$
$aaBB$	$\frac{1}{4}r^2$	$-a_A + a_B$	$aaBB$	$\frac{1}{4}(1-r)^2$	$-a_A + a_B$
$aaBb$	$\frac{1}{2}r(1-r)$	$-a_A + d_B$	$aaBb$	$\frac{1}{2}r(1-r)$	$-a_A + d_B$
$aabb$	$\frac{1}{4}(1-r)^2$	$-a_A - a_B$	$aabb$	$\frac{1}{4}r^2$	$-a_A - a_B$

The algebra is tedious, but it can be shown that the  $F_2$  variance is:

$$(\frac{1}{2})(a_A^2 + a_B^2) + \delta(1 - 2r)a_A a_B + \frac{1}{4}(d_A^2 + d_B^2) + \frac{1}{2}(1 - 2r)^2 d_A d_B$$

where  $\delta = +1$  for coupling and  $-1$  for repulsion.

On selfing, the frequency of homozygotes increases and the frequency of heterozygotes decreases until after many generations the latter approaches zero and we have a population of homozygotes,  $AABB$ ,  $AAbb$ ,  $aaBB$  and  $aabb$ . However, the rapid approach to homozygosity due to recurrent selfing prevents linkage equilibrium ever being attained (Kearsey 1985). Each generation, the selfing of the remaining heterozygotes releases more additive genetic variation, until at  $F_\infty$ , there are no heterozygotes left. It can be shown that starting from zero in the  $F_1$  generation, the additive genetic variance increases by the following amount each selfing generation ( $t$ ), i.e. between  $F_2$  individuals, within  $F_2$  progenies (i.e. within  $F_3$  families), within  $F_3$  progenies (i.e. within  $F_4$  families) etc.:

Between  $F_2$  individuals ( $t = 1$ ):

$$(\frac{1}{2})(a_A^2 + a_B^2) + 2\delta(\frac{1}{2})(1 - 2r)a_A a_B$$

The average within  $F_2$  progenies variance is the sum over progenies of each within progeny variance multiplied by its frequency. So let us consider the  $AB/ab$  selfing series. On selfing the  $F_2$ 's, the variation for  $AABB$ ,  $AAbb$ ,  $aaBB$  and  $aabb$  is zero; for both  $AaBB$  and  $Aabb$  it is  $\frac{1}{2}a_A^2$  (from  $\frac{1}{4}AA$ ,  $\frac{1}{2}Aa$  and  $\frac{1}{4}aa$ ) and for both  $AAb$  and  $aaBb$  it is  $\frac{1}{2}a_B^2$ , in each case multiplied by  $\frac{1}{2}r$

(continued)

**Box 13.2** (continued)

(1-r); for AB/ab it is  $(\frac{1}{2})(a_A^2 + a_B^2) + (1-2r)a_Aa_B$  with frequency  $\frac{1}{2}(1-r)^2$  and for Ab/aB it is  $(\frac{1}{2})(a_A^2 + a_B^2) - (1-2r)a_Aa_B$  with frequency  $\frac{1}{2}r^2$ . We can do likewise for the Ab/aB selfing series. The overall result is ( $t = 2$ ):

$$(\frac{1}{2})^2(a_A^2 + a_B^2) + 2\delta(\frac{1}{2})^2(1-2r)^2a_Aa_B$$

Each selfing generation, there are just four types of segregating family derived from Aa, Bb, AB/ab and Ab/aB heterozygotes with variances  $\frac{1}{2}a_A^2$ ,  $\frac{1}{2}a_B^2$ ,  $(\frac{1}{2})(a_A^2 + a_B^2) + (1-2r)a_Aa_B$  and  $(\frac{1}{2})(a_A^2 + a_B^2) - (1-2r)a_Aa_B$ . One just has to work out their frequencies in each generation, which is tedious; but a pattern soon emerges. The frequencies for F<sub>3</sub> progenies (F<sub>4</sub> families) in the AB/ab selfing series are  $r(1-r)(1-r+r^2)$  for both Aa and Bb,  $\frac{1}{4}(1-r)^4 + \frac{1}{4}r^4$  for AB/ab and  $\frac{1}{2}r^2(1-r)^2$  for Ab/aB. The overall result for the average within F<sub>3</sub> progenies (F<sub>4</sub> families) is ( $t = 3$ ):

$$(\frac{1}{2})^3(a_A^2 + a_B^2) + 2\delta(\frac{1}{2})^3(1-2r)^3a_Aa_B$$

Hence for average within F<sub>t</sub> progenies we have:

$$(\frac{1}{2})^t(a_A^2 + a_B^2) + 2\delta(\frac{1}{2})^t(1-2r)^ta_Aa_B.$$

Hence the additive variance at F<sub>∞</sub> is:

$$\begin{aligned} & \sum [(\frac{1}{2})^t(a_A^2 + a_B^2) + 2\delta(\frac{1}{2})^t(1-2r)^ta_Aa_B] \\ &= (a_A^2 + a_B^2)\sum (\frac{1}{2})^t + 2\delta a_Aa_B \sum (\frac{1}{2})^t(1-2r)^t, \end{aligned}$$

where summation is from  $t = 1$  to  $t = \infty$ .

Now if we consider summation from  $t = 1$  to  $n$ ,

$$\begin{aligned} \Sigma(\frac{1}{2})^t - \frac{1}{2}\Sigma(\frac{1}{2})^t &= \frac{1}{2} - (\frac{1}{2})^{n+1}, \text{ so that when } n \rightarrow \infty, \Sigma(\frac{1}{2})^t - \frac{1}{2}\Sigma(\frac{1}{2})^t \\ &= \frac{1}{2}\Sigma(\frac{1}{2})^t = \frac{1}{2} \end{aligned}$$

Therefore  $\Sigma(\frac{1}{2})^t = 1$ .

$$\begin{aligned} \sum (\frac{1}{2})^t(1-2r)^t &= \sum [(1-2r)/2]^t, \text{ and } \sum [(1-2r)/2]^t - [(1-2r)/2] \sum [(1-2r)/2]^t \\ &= [(1-2r)/2] - [(1-2r)/2]^{n+1}, \end{aligned}$$

so that when  $n \rightarrow \infty$ ,  $\Sigma[(1-2r)/2]^t - [(1-2r)/2]\Sigma[(1-2r)/2]^t$  ( $= \frac{1}{2}(1+2r)$ )  $\Sigma[(1-2r)/2]^t = (1-2r)/2$

(continued)

**Box 13.2** (continued)

Therefore  $\Sigma(\frac{1}{2})^l(1-2r)^l = (1-2r)/(1+2r)$

Hence the additive variance at  $F_\infty$  is:  $(a_A^2 + a_B^2) + 2\delta[(1-2r)/(1+2r)]a_Aa_B$

This can be summed over all loci and pairs of loci ( $i$  and  $j$ ) to give the total additive variance at  $F_\infty$ :

$$2V_A^\# = 2V_A^* + 2\Sigma\Sigma\delta[(1-2r)/(1+2r)]a_ia_j$$

Prediction will also be less precise in the presence of linkage. Some simple theory for two linked loci in the absence of epistasis is given in Box 13.2.

The total additive variance at  $F_\infty$  is:

$$2V_A^* + 2\Sigma\Sigma\delta[(1-2r)/(1+2r)]a_ia_j,$$

where  $\delta = +1$  for coupling and  $-1$  for repulsion linkages between loci  $i$  and  $j$ .

The  $F_3$  mean is unaffected by linkage (in the absence of epistasis), but twice the variance of  $F_3$  family means is:

$$2V_A^* + 2\Sigma\Sigma\delta(1-2r)a_ia_j + \frac{1}{2}V_D^* + \frac{1}{4}\Sigma\Sigma(1-2r)^2d_Ad_B.$$

If the predominant phase is coupling, in the absence of dominance, the predicted frequency of recombinant inbred lines exceeding the target will be over optimistic, whereas if the predominant phase is repulsion, they will be over pessimistic (i.e. conservative). Directional dominance will inflate the variance of family means. Kearsey (1985) has shown that if the  $F_2$  generation is random mated, as a result of recombination, the variance changes to:

$$V_A^* + \Sigma\Sigma\delta(1-2r)(1-r)a_ia_j + V_D^* + \frac{1}{2}\Sigma\Sigma(1-2r)(1-r)^2d_Ad_B.$$

If this generation is now selfed, twice the variance of the family means is a better predictor of the additive variance at  $F_\infty$  because  $(1-r)$  is a good approximation of  $1/(1+2r)$ :

$$2V_A^* + 2\Sigma\Sigma\delta(1-2r)(1-r)a_ia_j + \frac{1}{2}V_D^* + \frac{1}{4}\Sigma\Sigma(1-2r)(1-r)^2d_Ad_B.$$

A breeder may not wish to go to the trouble of random mating the  $F_2$  generation, but if he or she did so, would continue selfing from this generation. Under these circumstances, the total additive variance at  $F_\infty$  would also change; albeit with less bias from linkage.

Despite possible biases from linkage and epistasis, in *Nicotiana rustica*, Jinks and Pooni (1980) found that predictions based upon samples of  $F_3$  families could be used with the same confidence as those based on a sophisticated triple test cross analysis. Furthermore, the  $F_3$  families could be replicated over environments to predict the environmental sensitivity of genotypes, provided more than two environments were used to increase the precision of estimation. Kearsey and

Pooni (1996) also emphasized that many experiments in a wide range of crops have confirmed the robust reliability of this approach with  $F_3$  families. For example, Thomas (1987) recommended the use of random  $F_3$  families for cross prediction in spring barley, based on ranking six crosses for a number of traits. In contrast, in spring wheat, van Oeveren and Stam (1992) and van Oeveren (1993) concluded that such cross prediction was not effective because it was often erroneous due to severe bias on estimates of the genetic parameters. This bias was caused by non-additive genetic effects and differences in growing conditions between the  $F_3$  selection environment and that of the subsequent cultivar, especially with respect to inter-genotypic competition. Hence different breeders will sometimes reach genuinely different conclusions based on experimental results, and this emphasizes the importance of doing appropriate experiments, both at the start and during the course of long breeding programmes.

## Number of Crosses and Population Size of Progenies

So what should be the relative allocation of finite resources to the number of crosses and the population size of progenies? If a lot of information is known about potential parents, one can advocate making few crosses of large progeny size. Witcombe and Virk (2001) argued that this strategy is ideally suited to the particular constraints and advantages of participatory plant breeding. The aim is to make an improvement over the cultivar currently being grown by relatively poor farmers. Under these circumstances a popular cultivar can be crossed with genotypes with complementary traits from multi-year and multi-site evaluation trials. However, one may want to maintain the momentum of a breeding programme by using potential cultivars as parents in another cycle of crossing before they have completed extensive evaluation, including comparisons with potential cultivars from other programmes. Under these circumstances one would probably make a larger number of crosses with the consequence of smaller progeny sizes. If the variation between and within crosses is similar (in fact identical in absence of epistasis), one could argue that the number of crosses and progeny sizes should be roughly equal, and the considerations below about progeny size will then apply to the initial number of crosses. Subsequently, one could discard whole crosses, and perhaps even sow more seed of the most promising ones. But what should be the progeny size, and the number of crosses?

Simmonds (1996) pointed out that too-large progenies are wasteful because of the relationship between the intensity of selection and population size. From the properties of the normal distribution, the proportions of a population which are expected to exceed 1.28, 2.33, 3.09 and 3.72 standard deviations from the mean are 1 in 10, 100, 1000 and 10,000, respectively. Simmonds (1996) considered a sample size of 200 to be big enough as it would contain a top five with a mean greater than two standard deviations, and noted that greater extremes may be difficult and expensive to find anyway. Thus even with few crosses, population sizes of more than 1000 may not be justified. In a quantitative genetics analysis of ten crosses of

modern Indica cultivars of rice in Sri Lanka, Perera et al. (1998) concluded from  $F_3$  cross prediction that it should be possible to obtain useful transgressive segregants for the majority of characters with 100 SSD lines from each, but that for a minority it would be necessary to produce as many as 460 such lines. Fahim et al. (1998) later confirmed in two crosses that for the majority of traits, the  $F_3$  predictions were close to the observed proportions of superior single seed descent lines at  $F_6$ . But what have breeders actually done in practice?

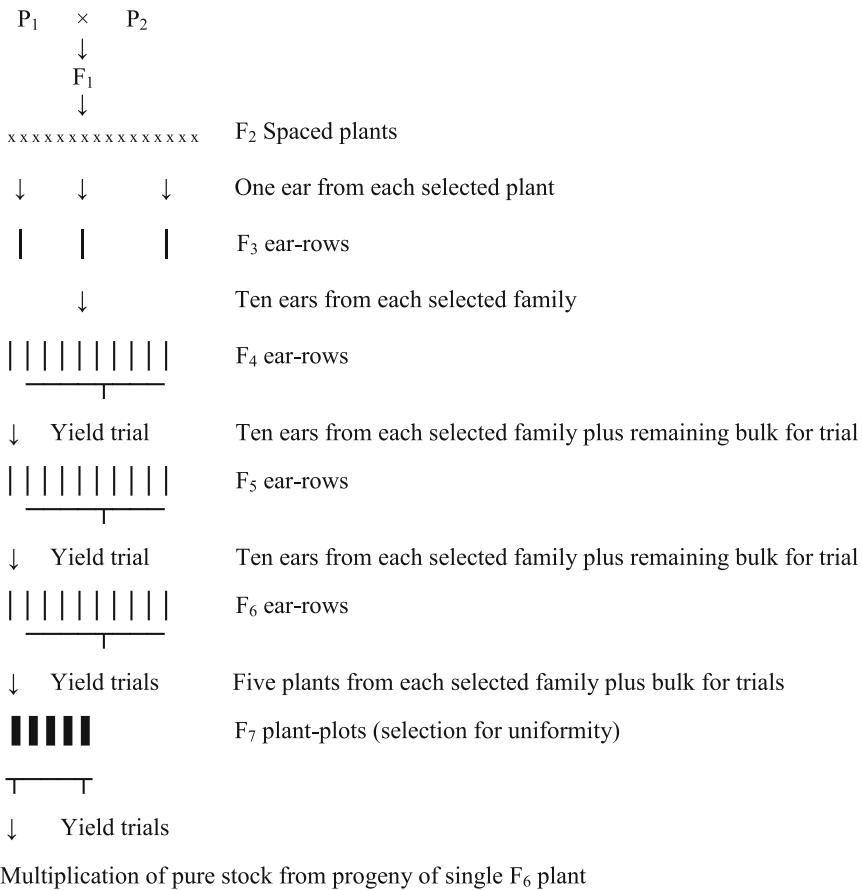
## Past and Current Practice

Witcombe and Virk (2001) conducted a postal survey in 1999 of European plant breeding companies and obtained information from 12 on a total of 19 field crops comprising wheat, barley, oilseed rape, peas and also potatoes, for which similar considerations apply as we saw in Chap. 11. The number of crosses at  $F_2$  per year ranged from 200 to 700, the size of  $F_2$  population from 150 to 4000 but with 1000 to 1500 typical, and the total size of the programme at  $F_2$  from 45,000 to 2.4 million. No low-cross-number (say 10 crosses) breeding programmes were reported. Likewise, they reported that high-cross-number strategies were pursued in the International Agricultural Research Centres such as CIMMYT (International Centre for Maize and Wheat Improvement) and IRRI (International Rice Research Institute). Hence high-cross-number strategies were the norm and still are, although precision breeding in future should lead to fewer crosses and appropriate population sizes or recurrent selection programmes for achieving objectives. However, it is perhaps worth reflecting on the fact that IR8, the first semi-dwarf rice of the green revolution, was selected at IRRI from 10,000  $F_2$  plants from the cross Peta  $\times$  Dee Geo Woo Gen, three quarters of which were discarded as tall, and then in the next generation from 298 pedigree rows (Peng and Khush 2003).

## Pedigree Inbreeding with Selection, Single Seed Descent, Doubled Haploids and Bulk Population Selection

### *Pedigree Inbreeding with Selection*

It is now time to examine the breeding process in more detail. It is instructive to start with Bingham's (Bingham 1979) very successful winter wheat breeding programme at the former Plant Breeding Institute in Cambridge, England, which was typical of pedigree inbreeding with selection. Each year he made 400–500 pair crosses (hand-pollinations) of cultivars or breeding lines with complementary characteristics in order to generate the genetic variability from which to select the rare plants with an improved combination of characters. The subsequent generations were then allowed to self-pollinate naturally. About 1500 plants of each cross were grown as spaced



**Fig. 13.2** Pedigree inbreeding with selection shown for one cross out of many

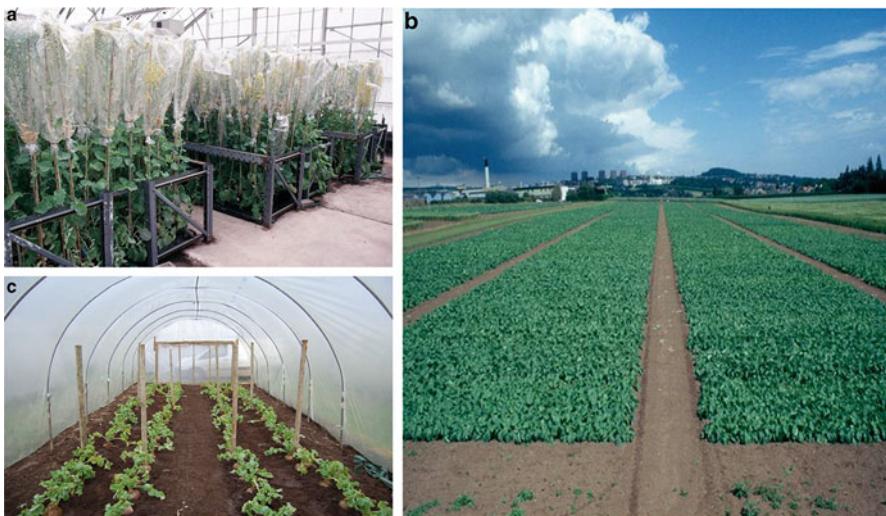
plants in the field at F<sub>2</sub>, making a total of 600,000–750,000. These were selected visually on a single plant basis for agronomic type, well filled ears and resistance to mildew, yellow rust, brown rust and *Septoria*, as well as for grain appearance and milling texture. A total of about 30,000 F<sub>2</sub> plants were grown on separately as ear to row observation plots in F<sub>3</sub> and selected more intensively. Sufficient grain was available for small scale milling and baking. Selected F<sub>3</sub> lines were represented by families of ten ear-rows in F<sub>4</sub> and the process of ear to row selection was repeated until the lines were stable (uniform and true-breeding). A uniform cultivar was typically multiplied from a single F<sub>5</sub> or F<sub>6</sub> plant. Yield trials began at F<sub>5</sub> at one site using a bulk of seed of the sister-lines of the one selected at F<sub>4</sub> (Fig. 13.2). More sites would be used at F<sub>6</sub> and F<sub>7</sub>. In the year prior to official National List Trials (NLT) (F<sub>7</sub> at the earliest), about 20 lines were still under consideration in yield trials at nine centres. These sites covered a range of areas, soil types and rotational entries, including first and fourth wheat crops. The most promising lines were entered for NLT at F<sub>8</sub>, at the earliest, and the first distribution of basic seed was made at F<sub>11</sub> at

the earliest, but more normally at  $F_{12}$ . With naturally self-pollinating crops such as wheat, the production of certified seed for sale to farmers is straightforward, but requires attention to detail, and careful monitoring to avoid poor quality seed and off-types in what should be a highly uniform crop.

Before considering the effectiveness of selection, we will look at the three main alternative breeding procedures that emerged during the twentieth century, namely single seed descent (SSD), the production of doubled haploids (DH) and bulk population selection.

### ***Single Seed Descent***

Single seed descent (SSD) was first proposed by Goulden in 1939 (Goulden 1939) and then developed by practical breeders in the 1960s and 1970s, along with theoretical considerations (Jensen 1988). The aim of an SSD programme is to produce advanced lines for assessment as potential cultivars, each of which is descended from a different  $F_2$  plant without selection. In crops like soybean and spring cereals (Sleper and Poehlman 2006), two or three generations can be grown each year in a glasshouse with large numbers of plants at really high density. In other crops, such as swedes (Bradshaw et al. 2009a), this is not possible because of vernalization requirements and the need to bag inflorescences to prevent cross-pollination (Fig. 13.3). Advanced lines can be the progenies of  $F_5$  or  $F_6$  plants, as in pedigree inbreeding. It is important that each advanced line is produced from a different  $F_2$  plant. If one simply used bulks to advance through the selfing



**Fig. 13.3** Swede breeding at SCRI (Dundee, Scotland) by single seed descent (funded by Limagrain UK Limited). (a) Production of 1597  $F_5$  plants by SSD from 1800  $F_2$  plants from 15 crosses in 1993. (b) Trial of 1037  $F_6$  families at SCRI in 1999. (c) Pre-basic seed production of Lomond (and Gowrie) in 2004 (grown commercially 2007)

generations before assessing single plant progenies as potential cultivars, a considerable loss of variability could occur. As explained by Kearsey and Pooni (1996), this is because individuals do not make equal contributions to the next generation and many will not be represented in later generations. Cober et al. (2009) reported that publicly developed soybean cultivars released in North America from 2004 to 2006 were overwhelmingly derived from biparental crosses (40/42) and single seed descent was the most popular breeding method (32/41). Soybean in the USA is a biotech crop, as we shall see later, and from 2001 to 2005 the US Patent and Trademark Office issued 388 cultivar patents, compared with four for the period from 1991 to 1995. There is concern that this may limit the use of patented cultivars as parents to the detriment of breeders in public sector and small or medium private institutions (Cober et al. 2009).

## Doubled Haploids

Guha and Maheshwari (1964) were the first to show that pollen of an angiosperm species, *Datura innoxia*, could be triggered into growth and that the resulting embryoids could be induced to develop into whole plants which were true haploids. As haploid plants have only one set of homologous chromosomes they cannot undergo normal meiosis and are sterile. However, their chromosome number can double spontaneously or be chemically induced to double, usually with colchicine. The resulting doubled haploid (DH) plants have two identical sets of chromosomes and hence are fertile and completely homozygous. Since 1964 efforts have been made to obtain DH plants in many plant species, both inbreeders and outbreeders; but currently efficient and reproducible protocols are available in only about 30 crop species (Szarejko 2012). These include inbreeding cereals such as wheat, rice and barley, as well as maize, potato, rapeseed and sugar beet. Anther culture is the most widely used technology (e.g. in rice, maize, wheat and potato) but isolated microspore culture is the method of choice when available (e.g. in barley and rapeseed) because of its high efficiency. The frequency of spontaneous chromosome doubling is high in barley and wheat. Wide crossing followed by chromosome elimination has been used in barley with *H. bulbosum*, in wheat with maize, and in potato with *S. tuberosum* Group Phureja as male parents; but anther and microspore culture are preferred in potato and barley, respectively. Wide crossing involves subsequent embryo rescue and treatment of plantlets with colchicine. Ovule culture is used in sugar beet, and again treatment of recovered plantlets with colchicine is required. In some fruit crops haploid production has been achieved via parthenogenesis induced by pollination with irradiated pollen. The pollen germinates, fails to fertilize, but stimulates haploid embryo development.

Doubled haploid (DH) lines in inbreeding crops are evaluated as potential cultivars. Those produced from an  $F_1$  are equivalent to homozygous lines ( $F_\infty$  lines) produced by SSD in the absence of linkage of genes governing the trait, and provided the method of haploid production has not directly affected the material.

The  $F_\infty$  lines from selfing may produce more transgressive segregation as a result of the additional opportunities for recombination between linked genes, DH lines being the result of one round of recombination. However, the short timescale to homozygous lines is a distinct advantage when DH lines can readily be produced. Furthermore, cycles of crossing and DH production are an attractive recurrent selection scheme.

### ***Bulk Population Selection***

Bulking during single seed descent is undesirable because of loss of variability, as already explained. However, the original bulk population procedure developed by Nilsson-Ehle in 1908 (Briggs and Knowles 1967) did have two advantages. Firstly, large populations of each cross could be grown in each generation, and secondly, these populations could be grown in environments where natural selection operated for desirable traits such as winter survival. In other words, the loss of variability was desirable. Breeders could also practise additional selection in the bulks if they wished. As a result bulk population breeding found its place as a simple method to operate and is still used today in various forms, as we shall see in a moment. The number of bulked generations before seeking pure lines would vary with the breeder and with the nature of the cross, but advanced lines could be the progenies of single  $F_5$  or  $F_6$  plants, as in pedigree inbreeding and single seed descent.

Bulking was taken further by Harlan and his associates who combined the  $F_2$  of several crosses into one composite population called a composite cross. For example, they bulked all 378 combinations of crosses between 28 cultivars of barley (Harlan and Martini 1929) and grew the composite without artificial selection to  $F_8$  when 2921 selections were made (Briggs and Knowles 1967). These proved comparable for yield to an equal number of selections obtained from the individual crosses by the pedigree method, thus demonstrating the merit of the approach. The experiment was continued for 29 generations by Suneson who concluded in 1956 that the yield gains achieved equalled those from the conventional and more costly (pedigree) breeding methods then in use (Suneson 1956). Almost 50 years later, in a review, Phillips and Wolfe (2005) concluded that composite cross populations may be an efficient way of providing heterogeneous crops and of selecting pure lines for low input farming systems characterized by unpredictable stress conditions. A low level of outcrossing would increase the genetic variation, but the structure of the heterogeneous crop would still be predominantly a mixture of highly inbred lines, albeit at different levels of inbreeding, as explained in Chap. 9.

Participatory rice breeding by the MASIPAG network in the Philippines provides a modern example of bulk population breeding (Medina 2012). The programme started in 1987 with the aim of providing new cultivars for indigenous organic farming systems and is conducted by some 67 farmers in

collaboration with researchers. It therefore provides an interesting contrast to the rice breeding of the ‘Green Revolution’. Bulk selection is practised because it is simple and convenient for the farmers to operate. For each cross, 10 F<sub>1</sub> plants are harvested and the seed bulked and used to raise a plot of 200 plants. During harvest early, medium and late maturing bulks are made for planting in three separate plots in the next cropping season. When these plots mature, short, medium and tall samples are taken to produce three sub-bulks of different heights within each maturity bulk, making nine combinations in total. In the ensuing generations, further sub-bulking may be done for other traits which the farmers consider to be important. This further sub-bulking soon creates many possible selections for the farmers to choose from to achieve their objectives. Each selection will be a collection of phenotypically similar inbred lines which are genetically different (diverse). There is no reason why these selections couldn’t be separated into their component lines by panicle-row selection. However, the farmers consider the diversity to be desirable for yield stability and durability of disease and pest resistance. Furthermore, segregating F<sub>3</sub> to F<sub>5</sub> lines are distributed to farmers in different locations so that they can select for local adaptability. For example, different farmer groups in different agro-climatic conditions made 12 selections with different characteristics from the cross between ‘Elon-elon’ and ‘Abrigo’. Further farmer assessment at several trial farms resulted in a reduction to four selections showing local adaptation to specific environments. After 24 years 1085 cultivars have been developed from 398 crosses (Medina 2012). In addition, another set of 508 cultivars from 328 crosses has been bred independently by farmers, without active support from researchers.

In Chap. 18 we will encounter a single backcross-selected bulk scheme for spring wheat which proved successful at CIMMYT in breeding for resistance to leaf and stripe rusts based on minor genes (Singh and Trethowan 2007). Susceptible but adapted cultivars provided 75 % of the genes in the first segregating generation.

### ***Best Method of Producing Inbred Lines***

So which of the methods just considered is the best one? The answer a breeder gives for a particular crop will be determined by the resources and timescale required to select superior true-breeding lines for multiplication as new cultivars. In short, the quicker and cheaper, the better. The simplicity of bulk population improvement makes it attractive for participatory plant breeding, as we have just seen. Single-seed descent is likely to be more attractive for spring than for winter cereals because the former do not have a vernalization period and hence the generation time is short. In contrast, the double-haploid method is likely to be preferred for rapidly developing inbred lines of winter cereals. Today with spring crops, shuttle breeding using off-season nurseries is common practice to achieve two generations in 1 year, and hence a generation time for pedigree inbreeding that is closer to single seed descent with its two or three generations a year. For example, with spring cereals,

New Zealand is commonly used for the off-season nurseries of breeding programmes in Northern Europe and North America. In addition to these practical considerations, it is also important to know the effectiveness of practising selection from  $F_2$  to  $F_5$  compared with delaying to  $F_6$  or selecting between doubled haploid lines. This is the topic for the rest of this chapter. However, it is important to point out that even if selection is delayed until inbred lines have been produced, their evaluation will still be a multi-trait, multistage process that involves independent culling levels; as already explained for vegetatively propagated clonal cultivars in Chap. 11.

## Selection Theory

### *Comparison of Doubled Haploids and Lines from Single Seed Descent*

Before considering the effect of selection during pedigree inbreeding, it is worth briefly mentioning the inbred lines that can be produced without selection as ( $F_1$ ) doubled haploids or from single seed descent (SSD).

The mean and variance of doubled haploid lines from the  $F_1AbC/aBc$  are derived in Box 13.3, those for SSD having been derived in Box 13.2, as done by Jinks and Pooni (1981b).

When there is linkage but no epistasis (all  $aa$ 's and  $dd$ 's equal zero):

$$\begin{aligned} P_1 &= AAbbCC = m + (a_A - a_B + a_C) = m + [a] \\ P_2 &= aaBBcc = m - (a_A - a_B + a_C) = m - [a] \\ F_1 &= AaBbCc = m + (d_A + d_B + d_C) = m + [d] \\ \text{Mean of doubled haploid lines and SSD lines} &= m \end{aligned}$$

Variance of doubled haploid lines =

$$\begin{aligned} a_A^2 + a_B^2 + a_C^2 - 2(1 - 2r_{AB})a_Aa_B + 2(1 - 2r_{AC})a_Aa_C - 2(1 - 2r_{BC})a_Ba_C \\ (= \text{estimate from } F_2 \text{ TTC}) = 2V_A^* + 2\sum\sum\delta(1 - 2r_{ij})a_i a_j \end{aligned}$$

Variance of SSD lines =  $2V_A^* + 2\sum\sum\delta[(1 - 2r_{ij})/(1 + 2r_{ij})]a_i a_j$

When there is epistasis but no linkage (all  $r$ 's equal a half):

$$\begin{aligned} P_1 &= AAbbCC = m + (a_A - a_B + a_C) + (-aa_{AB} + aa_{AC} - aa_{BC}) = m + [a] + [aa] \\ P_2 &= aaBBcc = m - (a_A - a_B + a_C) + (-aa_{AB} + aa_{AC} - aa_{BC}) = m - [a] + [aa] \\ F_1 &= AaBbCc = m + (d_A + d_B + d_C) + (dd_{AB} + dd_{AC} + dd_{BC}) = m + [d] + [dd] \end{aligned}$$

Mean of doubled haploid lines and SSD lines =  $m$

Variance of doubled haploid lines = Variance of SSD lines =

$$a_A^2 + a_B^2 + a_C^2 + aa_{AB}^2 + aa_{AC}^2 + aa_{BC}^2 = 2V_A^* + \Sigma\Sigma aa_{ij}^2$$

Hence the mean and variance of the inbred lines produced through doubled haploids and single seed descent are the same in the absence of linkage. In the presence of linkage, the variances but not the means differ, being greater for SSD when repulsion linkages predominate, and less for SSD when coupling linkages predominate. In the presence of linkage and epistasis, both the means and variances differ as shown in Box 13.3.

In going on to consider selection during pedigree inbreeding, we will assume no linkage and no epistasis for simplicity.

**Box 13.3: Doubled haploids from  $F_1 = AbC/aBc$  where three loci are linked with recombination frequencies  $r_{AB}$  between A and B,  $r_{AC}$  between A and C, and  $r_{BC}$  between B and C ( $r_{AC} = r_{AB} + r_{BC} - 2r_{AB}r_{BC}$ )**

Doubled haploid	Frequency	Genetic value
$AbC$	$\frac{1}{2}(1-r_{AB})(1-r_{BC})$	$a_A - a_B + a_C$ $-aa_{AB} + aa_{AC} - aa_{BC}$
$AbC$	$\frac{1}{2}(1-r_{AB})(1-r_{BC})$	$-a_A + a_B - a_C$ $-aa_{AB} + aa_{AC} - aa_{BC}$
$aBc$	$\frac{1}{2}r_{AB}(1-r_{BC})$	$a_A + a_B - a_C$ $+aa_{AB} - aa_{AC} - aa_{BC}$
$aBc$	$\frac{1}{2}r_{AB}(1-r_{BC})$	$-a_A - a_B + a_C$ $+aa_{AB} - aa_{AC} - aa_{BC}$
$Abc$	$\frac{1}{2}(1-r_{AB})r_{BC}$	$a_A - a_B - a_C$ $-aa_{AB} - aa_{AC} + aa_{BC}$
$Abc$	$\frac{1}{2}(1-r_{AB})r_{BC}$	$-a_A + a_B + a_C$ $-aa_{AB} - aa_{AC} + aa_{BC}$
$aBC$	$\frac{1}{2}r_{AB}r_{BC}$	$a_A + a_B + a_C$ $+aa_{AB} + aa_{AC} + aa_{BC}$
$aBC$	$\frac{1}{2}r_{AB}r_{BC}$	$-a_A - a_B - a_C$ $+aa_{AB} + aa_{AC} + aa_{BC}$

$$P_1 = AAbbCC = m + (a_A - a_B + a_C) + (-aa_{AB} + aa_{AC} - aa_{BC}) = m + [a] + [aa]$$

$$P_2 = aaBBcc = m - (a_A - a_B + a_C) + (-aa_{AB} + aa_{AC} - aa_{BC}) = m - [a] + [aa]$$

$$F_1 = AaBbCc = m + (d_A + d_B + d_C) + (dd_{AB} + dd_{AC} + dd_{BC}) = m + [d] + [dd]$$

$$\begin{aligned} \text{Mean of doubled haploid lines} &= m - (1 - 2r_{AB})aa_{AB} + (1 - 2r_{AC})aa_{AC} \\ &\quad - (1 - 2r_{BC})aa_{BC} = m + \Sigma\Sigma\delta(1 - 2r_{ij})aa_{ij} \end{aligned}$$

where  $\delta$  is +1 for coupling and -1 for repulsion and summation is over pairs of loci.

(continued)

**Box 13.3** (continued)

Mean of SSD lines (Jinks and Pooni 1981b) =

$$m + \Sigma \Sigma \delta [(1 - 2r_{ij}) / (1 + 2r_{ij})] aa_{ij}$$

Variance of doubled haploid lines =  $a_A^2 + a_B^2 + a_C^2 - 2(1 - 2r_{AB})a_A a_B + 2(1 - 2r_{AC})a_A a_C - 2(1 - 2r_{BC})a_B a_C + 4r_{AB}(1 - r_{AB})aa_{AB}^2 + 4r_{AC}(1 - r_{AC})aa_{AC}^2 + 4r_{BC}(1 - r_{BC})aa_{BC}^2 = 2V_A^* + 2\Sigma \Sigma \delta (1 - 2r_{ij})a_i a_j + 4\Sigma \Sigma r_{ij}(1 - r_{ij})aa_{ij}^2$

Variance of SSD lines (Jinks and Pooni 1981b) =  $2V_A^* + 2\Sigma \Sigma \delta [(1 - 2r_{ij}) / (1 + 2r_{ij})] a_i a_j + 8\Sigma \Sigma [r_{ij} / (1 + 2r_{ij})^2] aa_{ij}^2$

### ***Selection During Pedigree Inbreeding***

Kearsey and Pooni (1996) considered the applications of selfing theory in their book on Quantitative Genetics. Simple theoretical considerations led them to conclude that selection is least effective in the  $F_2$  unless the  $F_2$  individuals can be replicated vegetatively; otherwise, maximum advance can be achieved by selecting among homozygous ( $F_\infty$ ) lines. Selection among  $F_3$  or  $F_4$  families can also be moderately effective but only when family size is large and heritability is moderate to high. Cornish (1990a) presented a more sophisticated theory to describe the effects of a single round of selection during a selfing programme in terms of both the mean and the genetic variance of the inbred lines produced. He assumed constant environmental effects over generations, no genotype-environment interactions and individual plant assessment of the trait. For simplicity we will assume that the lines are completely homozygous inbreds ( $F_\infty$ ), although in practice they will probably be the progeny of  $F_5$  or  $F_6$  plants. The selection differential ( $S$ ) is the difference between the mean of the selected individuals or families and the generation mean. The response ( $R$ ) to selection is the difference between the mean of the inbreds produced by selection and the mean of the  $F_\infty$  generation produced without selection. Let us consider the response to selection in the  $F_2$  generation.

## **Response to Selection in the F<sub>2</sub> Generation**

We need to consider the regression of the F<sub>∞</sub> generation (inbred lines) on the F<sub>2</sub> generation (individual plants from which lines derived by single seed descent).

For a single locus we have:

F <sub>2</sub> genotypes	AA	Aa	aa	Mean
Frequency	1/4	1/2	1/4	
Genotypic value (x)	a	d	-a	1/2d
F <sub>∞</sub> lines from F <sub>2</sub> 's	AA	1/2AA, 1/2aa	aa	
Genotypic mean (y)	a	0	-a	0

The regression coefficient ( $b$ ) is the covariance of  $y$  and  $x$  divided by the variance of  $x$  ( $V_x$ ) which includes the plant to plant environmental variation  $V_E$ :

$b = 1/2a^2/(1/2a^2 + 1/4d^2 + V_E)$ , which is the narrow sense heritability ( $h_n^2$ ) for the F<sub>2</sub> generation.

Hence the response to selection is:

$R = bS = bi\sigma_P$ , where  $i$  is the intensity of selection and  $\sigma_P$  is the square root of the F<sub>2</sub> phenotypic variance ( $V_x = V_P$ ). Hence

$$R = i1/2a^2 / (1/2a^2 + 1/4d^2 + V_E)^{1/2}$$

Summing over loci we have:

$$R = iV_A^*/(V_A^* + V_D^* + V_E)^{1/2} = iV_A^*/V_P^{1/2}$$

where  $V_A^* = 1/2\Sigma a^2$ ,  $V_D^* = 1/4\Sigma d^2$ ,  $V_P = V_A^* + V_D^* + V_E$  and  $\Sigma = \text{summation over (unlinked) loci}$ , as defined by Kearsey and Pooni (1996). Furthermore,  $b = V_A^*/V_x = V_A^*/V_P$ .

Cornish (1990a) gave the formula for the response to a single round of family selection in any later generation. Toms et al. (1994a) extended this to sib selection, where for practical reasons the plants that are selfed are the sibs of the ones that are evaluated, and hence do not contribute to the estimate of their family mean. If family sizes are large, there is little difference between true family selection and sib selection as there is not much opportunity to exploit sampling (chance) variation in family means. Cornish (1990a) also quantified the reduction in variance as a result of selection. The reduction in variance is inherited by all subsequent selfed generations, and further rounds of selection reduce the genetic variation even further. Additional variation will however be created by further segregation during the selfing series. Here we will just consider the effect of selection in the F<sub>2</sub> generation.

## **Effect of Selection in the F<sub>2</sub> Generation on the Variance**

The regression of the F<sub>∞</sub> generation (inbred lines) on the F<sub>2</sub> generation (individual plants from which lines derived by single seed descent) can be expressed as:

$$y = a + bx + e$$

where  $a$  is a constant,  $e$  is the residual error and  $b$  the regression of  $y$  on  $x$ , i.e. the narrow-sense heritability of  $F_2$  individuals.

If the variances of  $y$ ,  $x$  and  $e$  are  $V_y$ ,  $V_x$  and  $V_e$ , and if  $x$  and  $e$  are uncorrelated, then the variance of the linear function  $a+bx+e$  is (Bulmer 1967):

$$\begin{aligned} V_y &= b^2V_x + V_e, \text{ so that} \\ V_e &= V_y - b^2V_x \end{aligned}$$

If selection in the  $F_2$  reduces the phenotypic variance of the selected individuals from  $V_x$  to  $(1-k)V_x$ , then the variance amongst the inbreds will become:

$$\begin{aligned} V_y' &= b^2(1-k)V_x + V_e \\ &= V_y - b^2kV_x \end{aligned}$$

The factor  $k$  depends on the intensity of selection. When selection is by truncation of a normal distribution (i.e. assume variation in the  $F_2$  generation displays a normal distribution):

$k = i(i-x)$  where  $i$  is the intensity of selection and  $x$  is the corresponding deviation of the point of truncation from the population mean (Falconer and Mackay 1996). Hence

$$V_y' = V_y - b^2i(i-x)V_x$$

Now  $b = V_A^*/V_x$ , so

$$\begin{aligned} V_y' &= V_y - bi(i-x)V_A^* \\ &= 2V_A^* - h_{n(F2)}^2i(i-x)V_A^*, \end{aligned}$$

where  $2V_A^*$  is the genetic variance between the inbred lines produced without selection.

Although we have shown the reduction in variance in the  $F_\infty$  generation, it occurs in all of the intermediate generations in the variance that can be attributed to segregation at gametogenesis in the  $F_1$ ; for example, the variance of  $F_3$  family means, the additive genetic variance of which is  $V_A^*$ . The remainder of the genetic variance of the inbreds ( $V_A^*$ ) results from further segregation from the  $F_2$  onwards and will be affected if the frequency of heterozygous loci is changed by selection. For example, when complete unidirectional dominance acts in same direction as selection, there is a tendency to select individuals and families which are slightly more heterozygous than would be expected in the absence of selection. Cornish (1990a) confirmed by computer simulation that this additional variance may also be affected by selection, but the assumption that it is unaffected is found to be

adequate. He also extended his approach, to the reduction in variance, to the selection of family means in later generations. Expected results were confirmed by computer simulation.

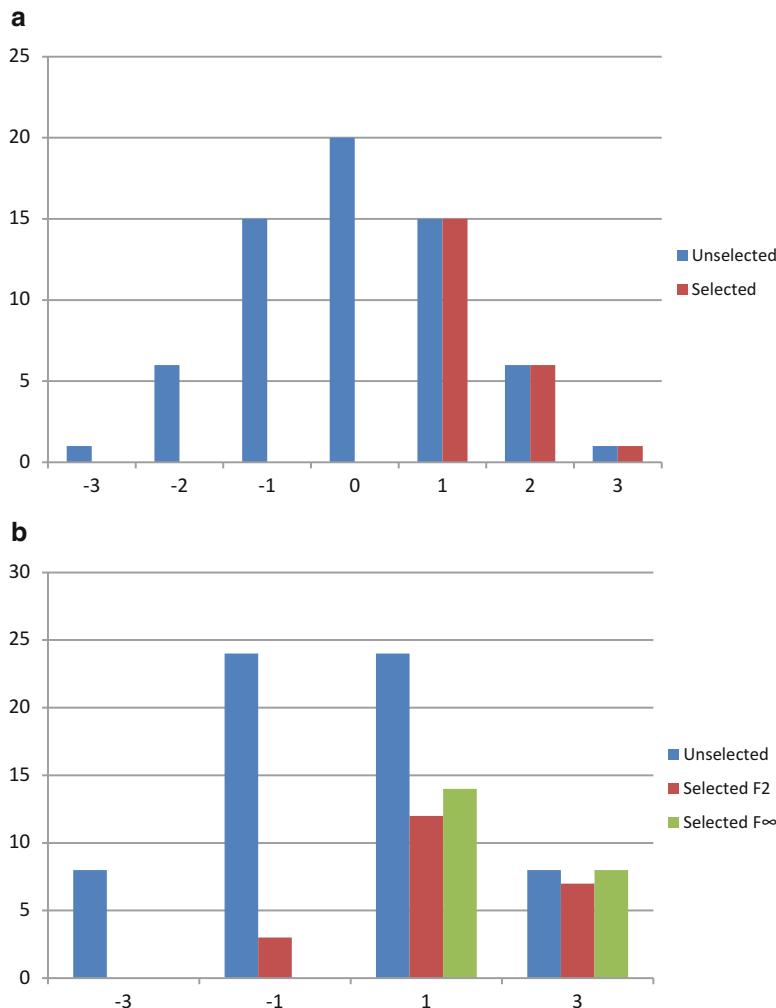
### ***Outcome of Selection at F<sub>2</sub> on the F<sub>∞</sub> Generation***

The outcome of this simple theory can be summarized as follows. Without selection, we produce an F<sub>∞</sub> generation with mean 0 (or  $m$  if we use  $m+a$ , etc.) and variance  $2V_A^*$ . With selection at F<sub>2</sub> we produce an F<sub>∞</sub> generation with mean  $R = iV_A^*/(V_A^* + V_D^* + V_E)^{1/2}$  and variance  $2V_A^* - h_n^2(F_2)i(i-x)V_A^*$ .

So does selection lead to a greater proportion of inbreds which exceed some target value? In other words, does the response to selection more than outweigh the effects of the reduction in the variance at F<sub>∞</sub>? This can be illustrated in a simple example with three unlinked loci segregating from the F<sub>1</sub>AaBbCc (Fig. 13.4). With a simple additive model and no environmental variation, if the best 22 F<sub>2</sub> individuals out of 64 are selected, one can see the increase in population mean (0 to 1.36) and reduction in variance (3 to 1.69) of the 22 derived lines at F<sub>∞</sub>. However, the best 22 lines selected at F<sub>∞</sub> are a slightly superior sample in terms of their mean (1.73 versus 1.36) and the proportion that have the maximum value of 3, namely 8/22 versus 7/22. In practice, the effectiveness of selection in the F<sub>2</sub> would be reduced by the environmental variation. With complete unidirectional dominance and loci of equal effect, there are fewer phenotypic classes and a skewed distribution at F<sub>2</sub> (Fig. 13.5). Again, if the best 27 F<sub>2</sub> individuals out of 64 are selected, one can see the increase in population mean (0 to 1.00) and reduction in variance (3 to 2.67) of the 27 derived lines at F<sub>∞</sub>. However, the best 27 lines selected at F<sub>∞</sub> are a superior sample in terms of their mean (1.59 versus 1.00) although the proportion with the maximum value of 3 is the same, namely 8/27. Having considered this specific example, we now need to look at more general results.

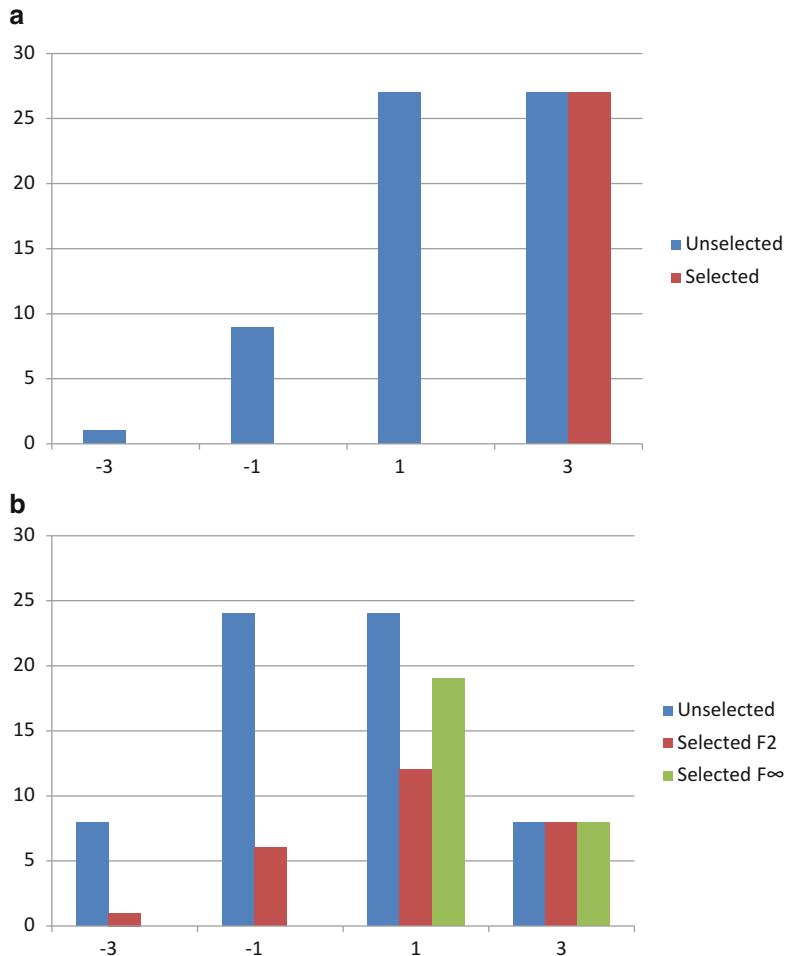
### ***Outcome of a Single Round of Selection***

The outcomes of the simple theory were explored by Cornish (1990a) by inserting numerical values in the deterministic equations and through computer simulations which included the complications of directional dominance and epistasis. Deviations in response due to directional dominance were relatively small, whereas complementary epistasis gave a better response for increasing expression than expected and duplicate epistasis a worse one because epistasis generates non-normality in the distribution of inbred lines. When a single round of selection was considered at one of the generations from F<sub>2</sub> to F<sub>6</sub>, the response was best when this was delayed to F<sub>6</sub>, and this can be considered to represent the optimum single seed descent programme. Experimental evidence for this result was provided by



**Fig. 13.4**  $F_2$  and  $F_\infty$  populations of size 64 from selfing  $F_1$  genotype  $AaBbCc$  (simple additive model with  $A=B=C=\frac{1}{2}$  and  $a=b=c=-\frac{1}{2}$ ). (a)  $F_2$  population of size 64 showing best 22 individuals. (b)  $F_\infty$  population of size 64 derived by single seed descent (SSD) from  $F_2$  population of size 64, together with best 22 lines selected at  $F_2$  and progressed to  $F_\infty$  by SSD and best 22 lines selected at  $F_\infty$ .

Jinks and Pooni (1981a) for final height in *Nicotiana rustica*. They concluded that with a single generation of selection, later was superior to early generation selection and family was superior to individual selection.



**Fig. 13.5**  $F_2$  and  $F_\infty$  populations of size 64 from selfing  $F_1$  genotype  $AaBbCc$  (simple additive-dominance model with  $A^- = B^- = C^- = 1$  and  $aa = bb = cc = -1$ ). (a)  $F_2$  population of size 64 showing best 27 individuals. (b)  $F_\infty$  population of size 64 derived by single seed descent (SSD) from  $F_2$  population of size 64, together with best 27 lines selected at  $F_2$  and progressed to  $F_\infty$  by SSD and best 27 lines selected at  $F_\infty$

### Selection in Two or More Generations

However, perhaps of more practical interest is a second paper of Cornish (1990b) in which he considered selection in two or more generations of a selfing series, with a fixed total number of plants from  $F_2$  to  $F_6$ . Computer simulations confirmed that only considerable skewness caused by epistasis, or a low number of segregating genes, led to serious departures from simple theoretical expectations. In all but the case with the highest heritability, the maximum response was obtained by applying selection in all five generations from  $F_2$  to  $F_6$ . These were between 20 and 25 %

better than those from  $F_6$  selection alone. However, the advantages of multi-stage designs over two-stage designs were not great, a result similar to that of Finney (1958) for multi-stage selection of genotypes (e.g. clones or inbred lines). These considerations imply that pedigree inbreeding with selection might be expected to be superior to SSD, but there is often insufficient experimental evidence for a definite conclusion in most if not all crops. In spring wheat, van Oeveren and Stam (1992) and van Oeveren (1993) concluded from computer simulations and experimental work that pedigree inbreeding with selection from  $F_3$  to  $F_5$  produces better lines than SSD, but it requires considerably more effort and resources for a relatively small genetic gain. In two crosses of Indica rice in Sri Lanka, Fahim et al. (1998) found that selection during selfing using pedigree, modified pedigree and bulk methods, produced lines at  $F_6$  which on average were superior to the 100 unselected ones produced by SSD; in other words selection was effective. However, none were significantly better than the best single seed descent line out of 100 for any character in either cross. They concluded that single seed descent is at least as effective as their other methods, is less costly, and more rapid when three generations can be raised each year.

### ***Correlated Responses to Selection***

Finally, a breeder may want to consider the consequences of selection on one trait for other traits. Toms et al. (1994b) extended the theory to predicting correlated responses to selection during selfing, and simulated pairs of characters correlated either by pleiotropy or linkage. Although comparison of expected with average observed correlated responses confirmed the accuracy of the theory, the observed sampling variance indicated that a wide variability of individual responses should be anticipated in practice.

### ***Value of Computer Simulations***

Although the implications that can be drawn from the previous considerations are limited by the simplifications adopted, they do illustrate an approach to designing efficient selfing programmes that comes from an understanding of responses to selection in terms of both means and variances. When simulations are coupled with experimental data from genetic analyses, including the effects of competition and genotype-environment interactions, they should be able to inform the design of actual breeding programmes. Methods of genetic analysis can be found in *The Genetic Analysis of Quantitative Traits* by Kearsey and Pooni (1996). Programme design will also be affected by considerations of the size of each generation that can be accommodated. Thus very large  $F_2$  populations of spaced plants may be possible, but environmental variation cannot be reduced by replication. Furthermore,

genotype-density interactions between spaced-plants at F<sub>2</sub> and normal spacing in subsequent generations may also be an issue. In contrast, rows of families can be replicated but the effectiveness of selection could still be reduced by inter-plot (genotypic) competition, particularly when single-row plots are used for families. Clearly the more experimental data that are available, the better will be the guidance from simulating different designs.

One other important conclusion from the simulations of Cornish (1990a, b) concerned the number of desirable alleles that were lost by chance during selfing, namely 4 out of the 20 on average. Hence when the number of genes segregating is large, the probability of obtaining the very best inbred, or even an individual approaching the best, is very small from a single cycle of breeding. As different elite inbred lines are likely to have fixed different desirable alleles, further progress can be made from elite line × elite line crosses, as breeders have done for the past 100 years. Hence the overall efficiency of a breeding programme must be judged over a number of cycles of crossing and selecting. This raises the issue of how soon a breeder has enough information on potential cultivars to use them as parents in the next cycle of crosses, as discussed in Chap. 11 on vegetatively propagated clonal cultivars. Eventually the genetic variation will run out, but as we saw in Chap. 10, there is no satisfactory way of predicting when this will happen.

## Conclusions on Breeding Methods

Plant breeders will no doubt continue to argue over the merits of pedigree inbreeding with selection, single seed descent, doubled haploids and bulk population selection. In truth, all four breeding methods are capable of producing superior inbred line cultivars. In many ways, the more difficult questions are the ones considered earlier, namely the size of programme required to guarantee success, the choice of parents and crosses to make, and the balance between number of crosses and size of progenies. Guidance was given but not definitive answers. As knowledge increases about the germplasm source of desirable alleles, their number and their chromosomal locations, it should become easier to design appropriate breeding programmes. No doubt these will often involve more than one cycle of crossing and selecting because of the number of alleles at different loci which must be combined and the need to break undesirable repulsion linkages. We will now conclude this chapter with a brief consideration of cultivar mixtures as an alternative to growing high yielding, genetically uniform, inbred-line cultivars in monoculture.

## Cultivar (Variety) Mixtures

Breeding success during the twentieth century has resulted in genetically variable landraces being replaced by potentially high yielding, genetically uniform, inbred-line cultivars; particularly in industrialized countries. The high and stable yields of such cultivars are achieved in practice through the input of synthetic fertilizers, herbicides, pesticides and fungicides, and irrigation. The desire to grow the best cultivar for a particular end-use leads to a few cultivars being grown over large areas of production (monocultures). The high inputs are not available in organic and low-input production systems and crops are exposed to more heterogeneous environments in both space and time. Under these conditions there is experimental evidence that higher and more stable yields can be achieved with genetically variable cultivars, as reviewed by Dawson and Goldringer (2012). One way to do this is through cultivar (variety) mixtures.

### Rice Example

A lot of research has been done on the use of mixtures for disease control, particularly for airborne fungal pathogens. One example will suffice here, with more given in Chap. 18 under “Resistance to Pests and Diseases”. Zhu et al. (2000) reported the results of a large study in Yunnan Province, China (Fig. 13.6), on using mixtures of commonly grown glutinous and hybrid rice (*Oryza sativa*) cultivars to control blast disease (*Magnaporthe grisea*). The disease spreads through multiple cycles of asexual conidiospore production during the cropping season, causing necrotic spots on leaves and necrosis of panicles. *M. grisea* exists as a mixture of pathogenic races that attack host genotypes with different resistance genes. Consequently in agricultural production, host resistance genes often remain effective for only a few years before succumbing to new pathogenic races. Single rows of susceptible glutinous rice (Huangkenuo or Zinuo) were dispersed between groups of four rows of resistant hybrid rice (Shanyuo22 or Shanyuo63), a common farmer practice to meet local demand for glutinous rice which is highly prized for speciality dishes. In the 2 years of the experiment the four mixtures were planted on 812 ha and 3342 ha areas, respectively. In the first year one fungicide spray was applied as a precaution, but none was needed in the second year. Mixtures were compared to monoculture control plots at 15 survey sites in the first year and at 30 survey sites in the second year. In the mixtures, the disease-susceptible cultivars had 89 % greater yield and blast was 94 % less severe than when they were grown in monoculture. Reduced disease severity certainly had a role in this yield response, but other factors such as improved light interception could also have had an influence. The mixed populations consistently produced more total grain per hectare than their corresponding monocultures. It was estimated that an average of 1.18 ha of monoculture crop land would need to be planted to provide the same



**Fig. 13.6** Rice paddy, Dali, Yunnan (6 June 2014)

amounts of hybrid and glutinous rice as were produced from 1 ha of a mixture. These findings increased farmer interest and the practice of mixtures expanded to more than 40,000 ha in the year 2000.

### ***Mixture Design***

When forming a mixture, genotypes with complementary resistance genes or mechanisms are sought together with adequate uniformity for agronomic traits, such as maturity for machine harvesting, and key aspects of product quality. Another important crop-specific decision is whether to make a physical mixture of seed before sowing, or to grow the components in different rows or strips. Such mixtures are usually created from pure line cultivars which have not been bred specifically for mixing ability or performance in heterogeneous populations. Under these circumstances, Jensen (1988) concluded from a review of research done during the twentieth century that the yield of the mixture was usually close to the mean of the components. The slight positive yield skew was not large, but Jensen thought that there was some merit in seeking the occasional mixtures that exceeded all their components in yield. Newton et al. (2008), for example, found that with seven winter barley cultivars in Scotland, the best improvement in yield compared to the monocultures was about 12 % without fungicide and 15 % with fungicide in the 7-component and 6-component mixtures, respectively. They developed methods of analysis to identify the beneficial or negative effects of component cultivars on malting quality traits, thereby providing data for optimising mixture design. The complex mixtures generally gave yield benefits without compromising quality. Furthermore, the performance of cultivars in simple mixtures such as 2-component ones was predictive of their performance in more complex mixtures.

### ***Breeding for Mixing Ability***

More recently, Dawson and Goldringer (2012) have advocated breeding specifically for mixing ability to produce mixtures that out-perform their best components. They point out that mixing ability of genotypes can be assessed in diallels (all combinations) of binary mixtures, but in practice, cultivar mixtures usually have 4–5 components. Hence selection of components from high-yielding bulk populations is an attractive proposition and worthy of more research.

# **Chapter 14**

## **Genetic Basis of Heterosis and Inbred Line Versus Hybrid Cultivars**

### **Introduction**

In Chaps. 2 and 12 we saw that hybrids were a big step forward in breeding cultivars of outbreeding crops that suffer inbreeding depression. The inbred lines of maize developed from open-pollinated cultivars by 5–7 generations of self-pollination were often weak and difficult to propagate. On crossing, however, vigour was restored (hybrid vigour) and the yield of the inbred line crosses (single crosses or  $F_1$  hybrids) usually exceeded that of the original open-pollinated cultivar from which the lines were developed, a phenomenon known as heterosis. Improvements in the inbred lines resulted in improvements in the hybrids, and we saw that single-cross hybrids eventually replaced double-cross ones. In other words, the hybrids remained superior to inbred lines for yield. We explained this in terms of the larger variance of inbred lines not compensating for their lower means and the large number of segregating loci that affect yield. The proportion of inbred lines that are superior to the mean of an open-pollinated cultivar decreases as the number of loci increases, whereas the proportion of superior single cross hybrids is always by definition one half. Hence  $F_1$  (single cross) hybrid breeding is an attractive proposition for outbreeding crops, particularly in the short term, whatever the genetic basis of heterosis.

The situation is different for crops that are tolerant of inbreeding and suffer little inbreeding depression. In the previous chapter we looked at how high yielding and uniform cultivars can be produced through crossing and inbreeding by various methods. It is now time to ask if  $F_1$  hybrids are the way forward with inbreeding crops, and to look at examples from rice, tomatoes and rapeseed. It is also time to look at the genetic basis of heterosis.

## The Genetic Basis of Heterosis

Today whenever possible F<sub>1</sub> hybrids are produced by crossing two highly inbred lines. When breeding inbred line cultivars the hybrid is the starting point of the inbreeding programme, whereas in hybrid breeding it is the end point of the programme; in other words, the new cultivar. Either way, interest centres on the circumstances under which the F<sub>1</sub> hybrid will out yield the better of its two parents (better-parent heterosis). Four explanations have been advanced over the years: a dispersion of dominant or partially dominant genes between the two parents; overdominance at individual loci; pseudo-overdominance as a result of tight linkage of two loci with desirable alleles in repulsion; and epistasis (Li et al. 2008). These explanations are not mutually exclusive so the problem is determining their relative importance in any particular hybrid, and this has proved difficult in practice. So despite a vast literature on the subject, the causes of heterosis are still being debated in conferences (Melchinger 2010) 100 years after the dominance (Davenport 1908; Bruce 1910; Jones 1917) and overdominance (East 1908; Shull 1908) hypotheses were first put forward. Jinks (1983) explained how the procedures of biometrical genetics can be used to define and analyze heterosis in terms of the actions and interactions of genes of the kinds that are well established in classical Mendelian genetics. We will start with a simple example of this approach before considering more recent publications involving molecular markers; but using the symbols in Kearsey and Pooni (1996).

### *A Simple Example*

Let us start by considering a simple example, the cross between AAbbCC and aaBBcc which results in the hybrid AaBbCc, where A, B and C are the alleles of increasing effect. Their genotypic values for an additive-dominance model with digenic interactions are as follows, using the F<sub>∞</sub> system where 'm' is the overall mean of all the inbred lines that can be derived from the cross (Kearsey and Pooni 1996).

$$\begin{aligned}P_1 &= AAbbCC = m + (a_A - a_B + a_C) + (-aa_{AB} + aa_{AC} - aa_{BC}) = m + [a] + [aa] \\P_2 &= aaBBcc = m - (a_A - a_B + a_C) + (-aa_{AB} + aa_{AC} - aa_{BC}) = m - [a] + [aa] \\F_1 &= AaBbCc = m + (d_A + d_B + d_C) + (dd_{AB} + dd_{AC} + dd_{BC}) = m + [d] + [dd]\end{aligned}$$

By definition the *a*'s are positive numbers whereas the *aa*'s, *d*'s and *dd*'s can be positive or negative depending on the direction of dominance and epistasis (*aa* and *dd* are symbols, not *a*'s and *d*'s multiplied together). The square brackets [] represent summation over loci and pairs of loci and hence are net effects. Dominance in the same direction is called unidirectional dominance and dominance in different directions is called ambidirectional dominance. As AAbbCC is the better parent

( $P_1$ ) with the larger score, we have  $m + [a]$ . The coefficient of  $[aa]$  is the square of the coefficient of  $[a]$ , and hence positive for both parents. Another way of expressing the net effects of  $[a]$  and  $[aa]$  is as follows:

$$[a] = r_a \times \Sigma a; (0 < r < 1)$$

$$[aa] = r_{aa} \times \Sigma aa; (-1 < r < 1)$$

Summation is over loci or pairs of loci and the  $r$ 's are coefficients of gene association/dispersion ( $r_a = r_{aa} = 1$  for complete association,  $r_a = 0$  and  $r_{aa} = -1$  for complete dispersion because a maximum of half the alleles of decreasing effect can be in the better parent, assuming for simplicity equal effect, whereas one could envisage an extreme situation where interactions between pairs of alleles occur between dispersed alleles but not between associated ones). Methods of estimating all of the relevant parameters for analysing heterosis from generation means can be found in the book by Kearsey and Pooni (1996), along with an explanation of the generations required. If a large random sample of recombinant inbred lines is available from a heterotic cross (cross 1), the highest and lowest scoring line for a given trait such as yield will have most of the alleles of increasing and decreasing effect, respectively. If we assume that the association is complete, analysis of  $P_H \times P_L$  (cross 2) can also be used to investigate the genetic basis of heterosis as  $r_a$  and  $r_{aa}$  will both have values of one (Jinks 1983). Furthermore,  $[a]_1/[a]_2$  will be an estimate of  $r_a$  and  $[aa]_1/[aa]_2$  will be an estimate of  $r_{aa}$ , and hence used to determine the contribution of dispersion to heterosis.

In the absence of non-allelic interactions, linkage has no effect on the mean of any generation, and hence has no effect on the magnitude of heterosis; but as we shall see, can complicate the analysis of its cause. Linkage of interacting pairs of genes has no effect on the means of non-segregating generations ( $P_1$ ,  $P_2$  and  $F_1$ ), and hence does not affect the specification of heterosis or its magnitude. Such linkage does, however, affect the means of segregating generations and hence biases the estimates of the components of heterosis. This leads to incorrect estimates of the relative contributions of the different kinds of gene action and interaction to heterosis. Jinks (1983) concluded that in general, linkage of interacting genes will result in overestimation of the contribution of dominance to heterosis and underestimation of non-allelic interactions. Hence in any analysis it is important to test the adequacy of the genetic models before drawing conclusions.

Another complication which we are not going to consider is genotype  $\times$  environment interaction. An account can be found in the review by Jinks (1983). If homozygous and heterozygous genotypes respond differently to environmental change, the magnitude of the heterosis will vary with the environment and this must be allowed for in the models, analyses and interpretations. Furthermore, if genotypes differ in their sensitivity to environmental change, this trait may display heterosis which can be analysed to determine its cause. Finally, maternal effects can be detected by comparing the means of reciprocal crosses and incorporated into the analysis of heterosis (Jinks 1983). Here we are going to concentrate on examining the main causes of heterosis.

### ***Dispersion of Dominant or Partially Dominant Genes***

In the absence of gene interaction and other complicating factors, better-parent heterosis ( $F_1 > P_1$ ) simply requires that:

$$m + [d] > m + [a]; \text{ that is } [d] > [a] \text{ or } \Sigma d > r_a \times \Sigma a.$$

Strictly speaking this is better-parent heterosis of the positive kind where the better-parent is the one with higher value. The reverse can be true for some traits, but we will stick with traits such as yield where higher is better. As long as the average dominance ratio ( $\Sigma d / \Sigma a$ ) is greater than the degree of gene dispersion ( $r_a$ ), heterosis will occur. Assuming a constant dominance ratio ( $d/a$ ) of 'f',  $\Sigma d = f \Sigma a$ , and there will be heterosis provided  $f \Sigma a > r_a \Sigma a$ , that is  $f > r_a$ . Therefore very little dominance is required at individual loci to produce quite considerable heterosis if the genes are dispersed in the two parents ( $r_a = 0.5$  when better parent has 25 % of undesirable alleles). Analysis of generation means does not allow us to determine the average dominance ratio, nor does it allow us to determine if overdominance ( $d > a$ ) occurs at any loci. Only if  $r_a = 1$ , when  $P_1$  has all of the increasing alleles, would it definitely be necessary to invoke overdominance to explain better-parent heterosis. In the absence of overdominance it should be possible to produce recombinant inbred lines that out-yield the  $F_1$  hybrid. Jinks (1983) summarized the results of 30 years work on two crosses in *Nicotiana rustica* which displayed better-parent heterosis for final height. They represented the extremes of simplicity ( $1 \times 5$ ) and complexity ( $2 \times 12$ ) in the control of heterosis. In the first, a simple additive-dominance model was adequate and heterosis could be explained by a high degree of dispersion ( $r_a = 0.18$ ) of partially dominant genes (mean dominance ratio 0.365). The second was the most heterotic of all that had been examined, and non-allelic interactions were present which did not conform to either classical complementary or duplicate interactions. Nevertheless, dispersion ( $r_a = 0.38$ ) of dominant increasing alleles (dominance ratio between 0.67 and 0.89), possibly linked in repulsion, was the main cause of the heterosis.

### ***Average Dominance Ratio at Individual Loci***

The average dominance ratio can be estimated from an analysis of appropriate generation variances. With a constant dominance ratio ( $d/a$ ) of 'f',  $\Sigma d^2 = (f)^2 \Sigma a^2$  and  $\Sigma d^2 / \Sigma a^2 = (f)^2$  so that  $(\Sigma d^2 / \Sigma a^2)^{1/2}$  equals  $d/a$ . When as is usual,  $(d/a)$  varies over loci,  $(\Sigma d^2 / \Sigma a^2)^{1/2}$  is the (weighted) average dominance ratio. It can be estimated as  $(4V_D^* / 2V_A^*)^{1/2}$ , where  $V_A^* = 1/2 \Sigma a^2$ ,  $V_D^* = 1/4 \Sigma d^2$  and  $\Sigma = \text{summation over (unlinked) loci}$ , as explained by Kearsey and Pooni (1996). The genetic variation present in the  $F_2$  generation produced by selfing the  $F_1$  ( $AaBbCc$ ) is  $1/2 \Sigma a^2 + 1/4 \Sigma d^2 [= 1/2 \Sigma (a_A^2 + a_B^2 + a_C^2) + 1/4 \Sigma (d_A^2 + d_B^2 + d_C^2)]$ , when the loci

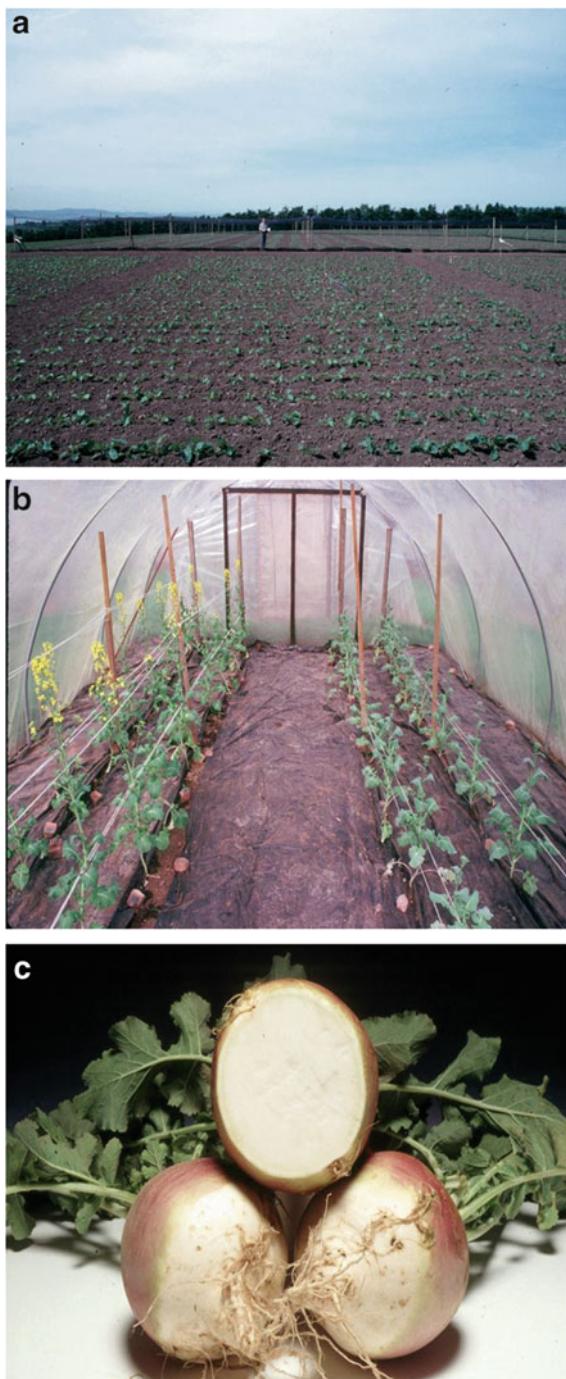
are unlinked. It is possible to test for and estimate  $\Sigma a^2$  and  $\Sigma d^2$  independently and with equal precision through a mating design known as the North Carolina Experiment III (NCIII) (Comstock and Robinson 1952). In this design, a number ( $n$ ) of  $F_2$  individuals ( $i$ ) are each crossed to the two parental lines  $P_1$  and  $P_2$  to generate pairs of families  $L_{1i}$  and  $L_{2i}$ , respectively. Commonly between 20 and 40  $F_2$ 's are used, and it is also common practice to use the unique  $F_2$  plants as the male parents because the inbred lines can be reproduced sexually and hence any one female need only be involved in one mating. It is in fact possible to create 'immortal'  $F_2$  populations by intercrossing recombinant inbred lines, as we shall see in a later example. The  $L_1$  and  $L_2$  families are assessed in replicated trials for the traits of interest, including yield. Appropriate analyses of variance (Kearsey and Pooni 1996) allow the true variation in  $L_{1i} + L_{2i}$  and  $L_{1i} - L_{2i}$  to be determined, these being estimates of  $1/2\Sigma a^2$  and  $1/2\Sigma d^2$ , respectively (Table 14.1 with  $r = 0$ ). In the analyses of variance, the components of variance are for  $(L_{1i} + L_{2i})/2$  and  $(L_{1i} - L_{2i})/2$  and hence are estimates of  $1/8\Sigma a^2$  and  $1/8\Sigma d^2$ , respectively, where the component for differences is one half of the traditional formulation of the interaction component of variance (parents  $\times$   $F_2$  individuals in  $2 \times n$  table of  $L$  family means). The analysis of variance also provides a variance ratio test for overdominance (interaction mean square divided by the mean square for  $F_2$  individuals). If recombinant inbred lines are produced from the  $F_2$  by single seed descent, these can also be backcrossed to the two parental lines  $P_1$  and  $P_2$  to generate pairs of families  $L_{1i}$  and  $L_{2i}$ , respectively; but now the true variation in  $L_{1i} + L_{2i}$  and  $L_{1i} - L_{2i}$  are estimates of  $\Sigma a^2$  and  $\Sigma d^2$ . Furthermore, the recombinant inbred lines provide a direct answer to the question of how easy it is to produce ones that out yield the  $F_1$ .

Kearsey and Pooni (1996) concluded that the experimental evidence suggested that dominance ratios measured as  $(4V_D^*/2V_A^*)^{1/2}$  are invariably  $\leq 1$ , despite the fact that most failures of the basic assumptions in the genetic models (no linkage and no epistasis) would tend to inflate the estimates. They also concluded that the size of the dominance ratio is in general a function of the trait rather than the breeding system of the species. The higher dominance ratios are typical of traits related to reproductive or competitive ability, such as yielding ability in crop plants, as found by Ramsay et al. (1994) for dry-weight yield in swedes (Fig. 14.1). The most important conclusion for plant breeding follows. When the heterosis is due to moderate dominance and allele dispersion, it should prove quite easy to produce transgressive inbred lines from the cross which perform better than the  $F_1$ . A good example is the swede cultivar Kenmore, an  $F_6$  line which out yielded the  $F_1$  hybrid (13.3 t/ha versus 12.3 t/ha; better parent 11.2 t/ha) from which it was produced by pedigree inbreeding with selection (Bradshaw and Wilson 1993). Furthermore, both Kenmore and the  $F_1$  out yielded seven commercial cultivars over 2 years. It was concluded that high yielding heterotic  $F_1$ 's should be the starting point rather than the finishing point of swede breeding programmes. Nevertheless, hybrid swedes remain attractive to commercial breeders who can maintain control of the parental lines while growers have to buy expensive seed for each new crop because the hybrids do not breed true.

**Table 14.1** North Carolina Experiment III: variation in  $L_1 + L_2$  and  $L_1 - L_2$  for two loci linked in coupling (a) and repulsion (b) with recombination frequency  $r$

$F_2$	Frequency	$\times AABB = L_1$	$\times aabb = L_2$	$L_1 + L_2 - d_A - d_B$	$L_1 - L_2 - a_A - a_B$
<b>(a) <math>AABB \times aabb \rightarrow AB/db \rightarrow F_2</math></b>					
$AABB$	$\frac{1}{4}(1-r)^2$	$d_A + a_B$	$d_A + d_B$	$a_A + a_B$	$-d_A - d_B$
$AABb$	$\frac{1}{2}r(1-r)$	$a_A + \frac{1}{2}(a_B + d_B)$	$d_A + \frac{1}{2}(-a_B + d_B)$	$a_A$	$-d_A$
$AAbb$	$\frac{1}{4}r^2$	$a_A + d_B$	$d_A - d_B$	$a_A - a_B$	$-d_A + d_B$
$AaBB$	$\frac{1}{2}r(1-r)$	$\frac{1}{2}(a_A + d_A) + a_B$	$\frac{1}{2}(-a_A + d_A) + d_B$	$a_B$	$-d_B$
$AB/db$	$\frac{1}{2}(1-r)^2$	$\frac{1}{2}(a_A + a_B + d_A + d_B)$	$\frac{1}{2}(-a_A - a_B + d_A + d_B)$		
$Ab/ab$	$\frac{1}{2}r^2$	$\frac{1}{2}(a_A + a_B + d_A + d_B)$	$\frac{1}{2}(-a_A - a_B + d_A + d_B)$		
$Aabb$	$\frac{1}{2}r(1-r)$	$\frac{1}{2}(a_A + d_A) + d_B$	$\frac{1}{2}(-a_A + d_A) - a_B$	$-a_B$	$d_B$
$aaBB$	$\frac{1}{4}r^2$	$d_A + a_B$	$-a_A + d_B$	$-a_A + a_B$	$d_A - d_B$
$aaBb$	$\frac{1}{2}r(1-r)$	$d_A + \frac{1}{2}(a_B + d_B)$	$-a_A + \frac{1}{2}(-a_B + d_B)$	$-a_A$	$d_A$
$aabb$	$\frac{1}{4}(1-r)^2$	$d_A + d_B$	$-a_A - a_B$	$-a_A - a_B$	$d_A + d_B$
<b>(b) <math>AAbb \times aaBB \rightarrow Ab/db \rightarrow F_2</math></b>					
$AAbb$	$\frac{1}{4}r^2$	$a_A + d_B$	$d_A + a_B$	$a_A + a_B$	$-d_A + d_B$
$AABb$	$\frac{1}{2}r(1-r)$	$a_A + \frac{1}{2}(-a_B + d_B)$	$d_A + \frac{1}{2}(a_B + d_B)$	$a_A$	$-d_A$
$AAbb$	$\frac{1}{4}(1-r)^2$	$a_A - a_B$	$d_A + d_B$	$a_A - a_B$	$-d_A - d_B$
$AaBB$	$\frac{1}{2}r(1-r)$	$\frac{1}{2}(a_A + d_A) + d_B$	$\frac{1}{2}(-a_A + d_A) + a_B$	$a_B$	$d_B$
$AB/db$	$\frac{1}{2}r^2$	$\frac{1}{2}(a_A - a_B + d_A + d_B)$	$\frac{1}{2}(-a_A + a_B + d_A + d_B)$		
$Ab/ab$	$\frac{1}{2}(1-r)^2$	$\frac{1}{2}(a_A - a_B + d_A + d_B)$	$\frac{1}{2}(-a_A + a_B + d_A + d_B)$		
$Aabb$	$\frac{1}{2}r(1-r)$	$\frac{1}{2}(a_A + d_A) - a_B$	$\frac{1}{2}(-a_A + d_A) + d_B$	$-a_B$	$-d_B$
$aaBB$	$\frac{1}{4}(1-r)^2$	$d_A + d_B$	$-a_A + a_B$	$-a_A + a_B$	$d_A + d_B$
$aaBb$	$\frac{1}{2}r(1-r)$	$d_A + \frac{1}{2}(-a_B + d_B)$	$-a_A + \frac{1}{2}(a_B + d_B)$	$-a_A$	$d_A$
$aabb$	$\frac{1}{4}r^2$	$d_A - a_B$	$-a_A + d_B$	$-a_A - a_B$	$d_A - d_B$

**Fig. 14.1** Genetic basis of heterosis in swedes. (a) Two triple test crosses in swede evaluated at SCRI (Dundee, Scotland) in 1988 and 1989: dominance ratio for dry-weight yield was 0.50 and 0.94, respectively (Ramsay et al. 1994). (b) F<sub>1</sub> hybrid production (self-incompatible Criffel × Magres) at SCRI in 1990. (c) Swede cultivar Kenmore, an F<sub>6</sub> line (Magres × Criffel) which outyielded the F<sub>1</sub> hybrid



The corollary of these considerations is as follows. When superior recombinant inbred lines are obtained from a heterotic cross, one can conclude that overdominance and its equivalent overepistasis (see below) are not the cause of the heterosis.

### **Linkage and Pseudo-Overdominance**

Let us now consider the genetic variance in the  $F_2$  when our three loci are linked ( $F_1$  is  $AbC/aBc$ ), with a recombination frequency ( $r$ ) of  $r_{AB}$  between  $A$  and  $B$ ,  $r_{AC}$  between  $A$  and  $C$ , and  $r_{BC}$  between  $B$  and  $C$ . It would be less confusing to use the symbol  $R$  rather  $r$ , as  $r_a$  and  $r_{aa}$  have been used above for the coefficients of gene association/dispersion, but inconsistent with earlier chapters.

The frequencies of the gametes from  $Ab/aB$  are  $Ab \frac{1}{2}(1-r_{AB})$ ,  $aB \frac{1}{2}(1-r_{AB})$ ,  $AB \frac{1}{2}r_{AB}$  and  $ab \frac{1}{2}r_{AB}$ ; and likewise for  $bC/Bc$ ; but for  $AC/ac$  they are  $AC \frac{1}{2}(1-r_{AC})$ ,  $ac \frac{1}{2}(1-r_{AC})$ ,  $Ac \frac{1}{2}r_{AC}$  and  $aC \frac{1}{2}r_{AC}$ . The algebra is tedious, but it can be shown that the  $F_2$  variance is:

$$\begin{aligned} & \frac{1}{2}(a_A^2 + a_B^2 + a_C^2) + [-(1 - 2r_{AB})a_Aa_B + (1 - 2r_{AC})a_Aa_C - (1 - 2r_{BC})a_Ba_C] \\ & + \frac{1}{4}(d_A^2 + d_B^2 + d_C^2) + \frac{1}{2}[(1 - 2r_{AB})^2 d_A d_B + (1 - 2r_{AC})^2 d_A d_C + (1 - 2r_{BC})^2 d_B d_C] \\ & = V_A^* + [-(1 - 2r_{AB})a_Aa_B + (1 - 2r_{AC})a_Aa_C - (1 - 2r_{BC})a_Ba_C] + V_D^* \\ & + \frac{1}{2}[(1 - 2r_{AB})^2 d_A d_B + (1 - 2r_{AC})^2 d_A d_C + (1 - 2r_{BC})^2 d_B d_C]. \end{aligned}$$

However, in the NCIII (see Table 14.1 for coupling and repulsion linkage of two loci) the variance of  $L_{1i} + L_{2i}$  is:

$$\begin{aligned} & \frac{1}{2}(a_A^2 + a_B^2 + a_C^2) + [-(1 - 2r_{AB})a_Aa_B + (1 - 2r_{AC})a_Aa_C - (1 - 2r_{BC})a_Ba_C] \\ & = V_A^* + [-(1 - 2r_{AB})a_Aa_B + (1 - 2r_{AC})a_Aa_C - (1 - 2r_{BC})a_Ba_C] \end{aligned}$$

and the variance of  $L_{1i} - L_{2i}$  is:

$$\begin{aligned} & \frac{1}{2}(d_A^2 + d_B^2 + d_C^2) + [(1 - 2r_{AB})d_A d_B + (1 - 2r_{AC})d_A d_C + (1 - 2r_{BC})d_B d_C] \\ & = 2V_D^* + [(1 - 2r_{AB})d_A d_B + (1 - 2r_{AC})d_A d_C + (1 - 2r_{BC})d_B d_C] \end{aligned}$$

This simple example allows us to make a number of points.

When the loci are unlinked ( $r = \frac{1}{2}$ ),  $1 - 2r$  equals zero, and the  $F_2$  variance is  $V_A^* + V_D^*$ , as expected, and the variances of  $L_{1i} + L_{2i}$  and  $L_{1i} - L_{2i}$  are  $V_A^*$  and  $2V_D^*$ , as expected.

All the increasing alleles can be in coupling ( $ABC/abc$ ), but they can never all be in repulsion. In our example,  $A$  and  $C$  are in coupling.

Coupling linkages ( $A$  and  $C$ ) inflate the additive variance whereas repulsion linkages ( $A$  and  $B$ ,  $B$  and  $C$ ) deflate it.

The dominance variance is not affected by linkage phase as such because the coefficients of the ' $d \times d$ ' terms are always positive. However, the  $d$ 's can be

**Table 14.2** Linkage ( $r = 0.5$  or  $0.0$  at all loci) and pseudo-overdominance in NCIII (three loci  $A$ ,  $B$  and  $C$  with alleles in coupling  $ABC/abc$  and repulsion  $AbC/aBc$  with unidirectional complete dominance,  $a$ 's and  $d$ 's all equal to 1.0, and ambidirectional complete dominance,  $a$ 's all equal to 1.0 but  $d_A = d_C = 1.0$  and  $d_B = -1.0$ ): bottom line is average dominance ( $d/a$ ) ratio

NCIII	$r = 0.5$	$r = 0.0$	$r = 0.0$	$r = 0.0$	$r = 0.0$
Linkage		Coupling	Coupling	Repulsion	Repulsion
Dominance		Unidirectional	Ambidirectional	Unidirectional	Ambidirectional
$V(L_1 + L_2)$	$1\frac{1}{2}$	$4\frac{1}{2}$	$4\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
$V(L_1 - L_2)$	$1\frac{1}{2}$	$4\frac{1}{2}$	$\frac{1}{2}$	$4\frac{1}{2}$	$\frac{1}{2}$
$V(L_1 - L_2)/V(L_1 + L_2)$	1	1	$1/9$	9	1
$[V(L_1 - L_2)/V(L_1 + L_2)]^{\frac{1}{2}}$	1	1	$1/3$	3	1

positive or negative so that with linked loci, the dominance variance is inflated by unidirectional dominance (+ × + or − × −) but deflated by ambidirectional dominance (+ × − or − × +).

The bias in estimating  $V_A^*$  and  $V_D^*$  is reduced by a number of generations of random mating, starting with the  $F_2$  generation. Hence changes in  $V_A^*$  and  $V_D^*$  with random mating are indicative of linkage disequilibrium as the cause.

The consequences for estimating the average dominance ratio  $(4V_D^*/2V_A^*)^{\frac{1}{2}}$  from the variances ( $V$ ) of  $L_{1i} + L_{2i}$  and  $L_{1i} - L_{2i}$  can be seen in Table 14.2. It is the combination of repulsion linkage and unidirectional dominance that creates pseudo-overdominance.

The general conclusions are explained by Kearsey and Pooni (1996). The effect of dispersion is always to reduce the observed additive variance relative to the observed dominance variance and hence cause the dominance ratio to be inflated. The extent to which this occurs depends on the actual gene distribution along the chromosome and it is most pronounced when the + and − alleles alternate in the sequence. Breeders commonly attempt to bring together dispersed sets of genes in the hope of producing a heterotic hybrid or transgressive segregants. Under these circumstances pseudo-overdominance may occur and be mistaken for true overdominance.

## Epistasis

In the presence of gene interaction, better parent positive heterosis ( $F_1 > P_1$ ) implies that:

$$m + [d] + [dd] > m + [a] + [aa] \text{ or } [d] + [dd] > [a] + [aa].$$

Hence epistatic situations in which  $[dd] > [aa]$  will increase heterosis. Complementary epistasis occurs when dominant increasing alleles at each of two loci have a

proportionately greater effect when they occur together, and is recognized by [d] and [dd] having the same sign. Perhaps surprisingly, Kearsey and Pooni (1996) report that duplicate rather than complementary epistasis has been found in heterotic crosses; in other words, [d] and [dd] have opposite signs. Hence the interaction opposes the dominance effect rather than reinforces it; and duplicate genes, gene dispersion and directional dominance are commonly found when heterosis is analyzed.

Although we are interested in the effect of epistasis on the  $F_1$  and  $P_1$  means, we do need to ask about its likely effect on estimates of the additive and dominance variances and hence the estimate of the average dominance ratio. Furthermore, estimates of  $\Sigma aa^2$  ( $4V_{AA}^*$ ) and  $\Sigma dd^2$  ( $16V_{DD}^*$ ) would provide  $2(V_{DD}^*/V_{AA}^*)^{1/2}$  as an estimate of the (weighted) average value  $dd/aa$ . If  $dd/aa=f$  at all loci,  $\Sigma dd^2=(f)^2\Sigma aa^2$  and  $\Sigma dd^2/\Sigma aa^2=(f)^2$  so that  $(\Sigma dd^2/\Sigma aa^2)^{1/2}$  equals  $dd/aa$ . This estimate would complement  $(4V_D^*/2V_A^*)^{1/2}$  in determining the cause of heterosis. Unfortunately there is no breeding design or biometrical genetical analysis that will allow us to estimate  $V_A^*$ ,  $V_D^*$ ,  $V_{AA}^*$  and  $V_{DD}^*$  when all are present simultaneously. Epistasis is in fact difficult to detect in variances. In general, the interaction components of variance appear only as biases on estimates of  $V_A^*$  and  $V_D^*$ , as explained by Jinks (1983) for the estimates from the NCIII and Triple Test Cross (TTC) mating designs. He points out that the estimate of  $V_D^*$  from these designs, but not generally, is always inflated by non-allelic interactions irrespective of whether the interacting alleles are associated or dispersed in the parents, and whether they display complementary or duplicate interactions. In contrast, estimates of  $V_A^*$  may be inflated or deflated (most likely with dispersed duplicate genes) by the interaction bias. Nevertheless, with the exception of dispersed duplicate genes, the dominance ratio is unlikely to be spuriously inflated by the presence of non-allelic interactions and the bias in the ratio is relatively small.

One of the most commonly used tests for epistasis comes from the Triple Test Cross (TTC), which differs from the NCIII design in that the  $F_2$  individuals are not only crossed to  $P_1$  and  $P_2$ , but also to the  $F_1$  to give triplets of families  $L_1$ ,  $L_2$  and  $L_3$  (Kearsey and Jinks 1968). When the additive-dominance model is adequate,  $(L_1+L_2)-2L_3$  is zero for every  $F_2$  individual. Otherwise, the overall difference is due to additive by additive epistasis and the variation over  $F_2$  individuals is due to additive by dominance and dominance by dominance epistasis. This test has shown that in many plants the variance due to epistasis is often very small compared with its directional effect on the means. This may either be because its magnitude is small, or being an interaction of the main effects, it is not a major source of variation (Kearsey and Pooni 1996).

## Recent Examples of Research on the Genetic Basis of Heterosis

Since Jinks (1983) wrote his review of how the procedures of biometrical genetics can be used to define and analyze heterosis, the results of many experimental studies have been published. Furthermore, the advent of dense molecular marker maps has meant that attempts have been made to dissect heterosis into the effects of QTLs and pairs of QTLs linked to markers across the entire genome. Thus in our table for the NCIII design (Table 14.1), if we genotype the  $F_2$  population with molecular markers we will know the genotype of the  $F_2$  parent of the corresponding  $L_1$  and  $L_2$  families. Now consider a QTL ( $Q$ ) linked to a marker ( $M$ ) so that the cross becomes  $MMQQ \times mmqq$  or  $MMqq \times mmQQ$ . If the families are classified by the marker genotype of their parents,  $\frac{1}{2}(MM-mm)$  for  $L_1+L_2$  is an estimate of either plus or minus  $(1-2r)a_Q$ ,  $L_1-L_2$  is an estimate of minus  $(1-2r)d_Q$ , and their ratio is an estimate of  $d_Q/a_Q$ . Cockerham and Zeng (1996) have provided the detailed theory for the analysis and estimation of effects such as  $a_Q$  and  $d_Q$ . Zeng et al. (2005) have explained the advantages of the  $F_2$  genetic model over the  $F_\infty$  one used in this book, namely the former is an orthogonal model. Using different models does not influence the detection of QTL and epistasis but it does influence the estimation and interpretation of genetic effects.

For example, in the  $F_2$  model, the genotypic values are measured relative to the  $F_2$  mean so that  $AA$ ,  $Aa$  and  $aa$  have genotypic values  $a-\frac{1}{2}d$ ,  $\frac{1}{2}d$ , and  $-a-\frac{1}{2}d$  rather than  $a$ ,  $d$  and  $-a$ . In both models the difference between homozygotes is  $2a$  and the  $F_1$  minus the midparent (MP) value is  $d$ . With the inclusion of digenic interactions in the  $F_2$  model and summation over loci and pairs of loci ( $d_i a_j \neq a_j d_i$ ), so both included in summation over pairs of loci represented as  $[ad]$ ), the difference between homozygotes is now  $2\sum a - \sum da$ , so that half the difference is  $\sum a - \frac{1}{2}\sum da$  ( $[a] - \frac{1}{2}[ad]$ ). Furthermore, the midparent heterosis ( $F_1-MP$ ) becomes  $[d]-[aa]$  instead of  $[d]-[aa]+[dd]$ , where  $[d]-[aa]=\sum d - \frac{1}{2}\sum aa$  ( $a_i a_j = a_j a_i$ , so only one included in  $[aa]$ ). Hence the additive and dominance effects are defined differently in the two models when there are two or more loci with epistasis. We shall see below that  $a-\frac{1}{2}\sum da$  and  $d-\frac{1}{2}\sum aa$  are defined as augmented additive and dominance effects in the  $F_2$  model. Hence it is important to be aware of the model being used in any particular scientific paper.

Another point worth making is as follows. If recombinant inbred lines are produced and used instead of the  $F_2$  population in a NCIII or TTC design with molecular markers, it should be possible to determine if the extreme phenotypes for a particular trait also have all of the QTL alleles of increasing and decreasing effect, respectively.

We will now concentrate on some recent results for two inbreeding crops, rice and tomato, where a major effort has gone into  $F_1$  hybrid breeding; and compare them with those from outbreeding maize.

## ***Maize***

We can take the investigation of Frascaroli et al. (2007) as a fairly typical example for maize. They analyzed heterosis in the single cross B73 × H99, where B73 belongs to the Iowa Stiff Stalk Synthetic (BSSS) heterotic group and H99, developed from Illinois Synthetic 60C, belongs to the Lancaster Sure Crop (LSC) group. The 142 recombinant inbred lines (RILs) obtained by single-seed descent from the original single cross were used as the female parents in a triple test cross (TTC), and all of the material was assessed for yield and related traits. Both classical biometrical genetical and QTL analyses were performed on the data. Here we will look at the results for grain yield, the trait that displayed the greatest heterosis. Both parents were relatively low yielding at 3.81 and 2.48 t/ha for B73 and H99, respectively. In contrast, their hybrid yielded 10.69 t/ha and thus displayed considerable better parent heterosis. No epistasis was detected in the TTC analysis and the estimates of  $V_A^*$  and  $V_D^*$  were 1.02 and 0.71, respectively, giving an average degree of dominance of 1.18 and providing evidence for overdominance. The QTL analysis revealed 21 main-effect QTLs for grain yield, with just four pairs of QTLs involved in digenic interactions: three ‘additive × additive’ interactions and one ‘dominance × dominance’ interaction. Hence the interactions were less important than the main effects. The QTL alleles for increasing yield were dispersed between the parents with a net effect in favour of B73. Sixteen of the 21 QTLs displayed overdominance, all in the direction of increasing yield. Hence the heterosis could be explained by a dispersion of alleles for increasing yield between the parents and directional overdominance for increasing yield. It was not possible to distinguish between true overdominance and pseudo-overdominance. This would require determining the effect on the estimates of  $V_A^*$  and  $V_D^*$  of a number of rounds of random mating after selfing the hybrid. Some authors also feel that a way forward is comparison of inbred lines and their hybrids by transcription profiling using microarray technology.

## ***Rice***

We can take as our examples for rice, three studies connected by common parents. Although some of the genetical analyses are even more sophisticated than the one for maize, they encounter difficulties in explaining the observed heterosis.

Zhou et al. (2012) analyzed yield heterosis in an elite indica rice hybrid Shanyou 63 that has been widely grown in China. An immortalized F<sub>2</sub> population was constructed comprising 278 crosses from 210 recombinant inbred lines derived by single seed descent from the cross between Zhenshan 97 and Minghui 63, the parents of Shanyou 63. The population was assessed for yield, number of tillers per plant, number of grains per panicle and grain weight during 1998 and 1999 in replicated field trials on the experimental farm of Huazhong Agricultural

University, Wuhan, China. An ultra-high-density SNP map of the population consisting of 1619 bins was constructed by sequencing the 210 recombinant inbred lines. There were three genotypes in each bin, a homozygote for each parental genotype and their heterozygote. Single-locus and epistatic genetic effects were calculated for bins and pairs of bins across the whole genome and their contributions to the heterosis were estimated. As the effects detected in adjacent bins might result from a single bin, heterotic bins were merged into clusters according to their physical positions and similar effects. For yield, net positive overdominance was the most important contributor to heterosis, followed by net positive dominance  $\times$  dominance (DD) interactions. Overdominance also contributed the most to grains per panicle. DD interactions were more important than overdominance for tillers per plant, but the heterotic effects of this trait were not consistent between years. For grain weight, the contribution of overdominance was also greater than that for DD interactions and dominance, although the total amount of heterotic effect was relatively small. However, complementarities of the parents in gene expression and genomic sequences (insertions and deletions) suggested that pseudo-overdominance rather than true overdominance could be the main contributor to the observed heterosis. This emphasizes the continuing difficulties encountered when trying to determine the genetic basis of heterosis.

Li et al. (2008) took a different approach to analyzing heterosis in the hybrid between Zhenshan97 (indica) and Minghui63 and in another highly heterotic hybrid between 9024 (indica) and LH422 (japonica). They combined QTL analyses using molecular markers with NCIII analyses in which recombinant inbred lines (194 and 222, respectively) derived from the  $F_2$  populations were backcrossed to the parents. While it is not possible to do their extensive results full justice in a brief summary, the following points can be emphasized. Grain yield showed the strongest heterosis among the nine traits studied. In the first hybrid the proportion of QTL exhibiting overdominance was more than the proportion showing partial to complete dominance. In the second subspecies hybrid the reverse was true. A large number of digenic interactions were found, most between complementary loci with no detectable main effects. However, the study was not able to distinguish true overdominance from pseudo-overdominance.

Garcia et al. (2008) extended the NCIII analysis of Cockerham and Zeng (1996) to Multiple-Interval-Mapping (MIM) in which they estimated the augmented additive ( $a_r - \frac{1}{2}\sum da_{rs}$ ) and dominance ( $d_r - \frac{1}{2}\sum aa_{rs}$ ) effects of each QTL ( $r$ ) as well as epistatic effects between two QTLs ( $r$  and  $s$ ). The augmented additive and dominance effects are the net contributions of the QTL to parental difference and mid-parent heterosis, respectively, and are useful because the main effects contain epistasis that cannot be removed or estimated separately (recall Box 13.1). Garcia et al. (2008) also used the hybrid between 9024 (indica) and LH422 (japonica) for their analysis of heterosis, in which they backcrossed 194 recombinant inbred lines to the two parents. Again it is not possible to do their extensive results full justice in a brief summary. Six QTL for yield were detected. The augmented additive effect was positive for one QTL and negative for the other five, showing that the favourable alleles are distributed between the parents but with most in the indica

parent. The augmented dominance effects were not significantly greater than the augmented additive effects for any QTL, and both positive and negative effects occurred, indicating that the heterozygote was not always superior in the direction of the favourable allele. Hence there was both a lack of overdominance and directional dominance which led the authors to suggest epistasis as the cause of the heterosis. However, the results did not seem to entirely agree with those of Li et al. (2008), again emphasizing the continuing difficulties encountered when trying to determine the genetic basis of heterosis.

## **Tomato**

Semel et al. (2006) used a population of 76 introgression lines (ILs) to assess the contribution of overdominance to heterosis in tomato in the absence of epistasis. In each line a single marker-defined chromosome segment from the distantly related wild species *Solanum pennellii* replaced the homologous segment of the cultivated tomato (*S. lycopersicum*) variety M82. The set of lines represented the entire *S. pennellii* genome in a cultivated background. Each line was crossed with M82 to create 76 hybrids (ILHs). The lines, their hybrids, and M82 were assessed for 35 traits of which 11 were considered aspects of reproductive fitness. Each IL and ILH was compared with M82 to see if its chromosome segment had a QTL that affected the trait; and if so, the direction and magnitude of dominance. In total 841 QTLs were identified. Specific genomic regions with overdominance effects were overwhelmingly associated with yield-related phenotypes that defined traits for increased reproductive fitness. The authors argued that these effects were true overdominance and not pseudo-overdominance because dominant and recessive QTL were detected for all phenotypic categories but overdominance only for reproductive traits. They also argued that this association was the result of natural selection during evolution and was then domesticated by humans to improve the yields of crop plants. In other words, their results are relevant to breeding F<sub>1</sub> hybrid tomatoes.

## **Examples of Hybrid Breeding**

Hybrid breeding in inbreeding and partially inbreeding crops has progressed despite a lack of a complete understanding of the genetical basis of heterosis. Hybrid cultivars have seemed an attractive proposition whenever breeders have found commercially acceptable levels of better-parent heterosis and a method of hybrid seed production. In practice this has meant a yield advantage of around 20 % over existing conventional inbred line cultivars. We will finish this chapter by looking at grain sorghum, rice, tomatoes, rapeseed and sunflowers as examples.

## Grain Sorghum

One of the first examples of hybrid breeding in an inbreeding crop was grain sorghum (*Sorghum bicolor*) in the USA, made possible by the early discovery (Stephens and Holland 1954) of the essential components of the three-line system of hybrid seed production using genetic-cytoplasmic male sterility. One line (A) is the female parent of the hybrid as a result of genetic-cytoplasmic male sterility, the second line (B) is responsible for maintaining the sterility, and the third line (R) is the male parent of the hybrid that restores its fertility through a dominant restorer gene(s). In other words, once a source of CMS has been found, all male-fertile germplasm is either a maintainer or a restorer with respect to this source, as shown back in Chap. 12 in Fig. 12.5 for maize. The following information is taken from the chapter on breeding sorghum in *Breeding Field Crops Fifth Edition* by Sleper and Poehlman (2006).

Sorghum is the world's fifth leading cereal grain after maize, rice, wheat and barley. It is of African origin with its centre of diversity in Ethiopia, where it is adapted to flowering in short days. It is considered an inbreeding crop, although natural cross-pollination by wind averages about 6 %. It is tolerant of heat and drought and hence widely grown in the semiarid regions of sub-Saharan Africa, in the dry central parts of India, and in the parts of the Great Plains of the USA that are too hot and dry for growing maize. The grain is used for both animal feed and human food. The grain sorghums that were originally introduced into the USA in the middle of the nineteenth century came from different geographical areas; for example, the milo type from east-central Africa and the kafirs from South Africa. Many of the first cultivars originated by selection of off-type plants in introduced sorghum populations. Key mutants were those for dwarfness (four independent recessive genes) and early maturity (six major genes), which allowed mechanical harvesting and the extension of cultivation northwards. Suitable cultivars came from hybridizations followed by pedigree inbreeding. The hybridizations also revealed the vigour and yield of F<sub>1</sub> hybrids. Whilst selections from kafir × milo crosses combined the desirable features of each parent, milo × kafir crosses resulted in cytoplasmic male sterility. Thus milo germplasm provided the source of male sterile cytoplasm (A1) and a restorer gene ( $M_{SC1}$ ), and kafir germplasm could be used as a maintainer. In practice new parent lines are first evaluated for use as seed-parent lines (A and B) or as pollen-parent lines (R). In addition to  $M_{SC1}$ , pollen parents often require additional modifying genes to obtain good seed production in a wide range of environments. Thus hybrids replaced inbred line cultivars starting in 1956, and breeding passed from the public to the private sector. To date five major groups of cytoplasmic male sterility have been found (A1 to A5). In commercial production of hybrid seed, an isolated field (200 m from other sorghum) usually comprises 12 rows of the male-sterile A-line to four rows of the pollinator R-line. Today worldwide, cultivars are either inbreds or hybrids in the underdeveloped countries but F<sub>1</sub> hybrids in the more developed ones.

The hybrid sorghum breeding programme at the Texas Agricultural Experiment Station has been explained by Rooney (2012). Improvement of the A/B lines and R lines is done separately by hybridizations (at least one elite parent) followed by pedigree inbreeding with selection until the F<sub>5</sub> generation. The evaluations are done in contrasting environments to allow selection of widely adapted germplasm. The new B-lines are male-sterilized (converted to new A-lines) by using them as the recurrent male parent for five backcrosses, starting with an appropriate male-sterile A-line as female parent. Winter nurseries are used to reduce the time required for the process. Each generation, plants and progenies are selected for full male sterility and similarity to the B-line. Testcrossing begins at F<sub>5</sub> for R-lines and at BC<sub>3</sub> for the A/B-lines. New R-lines are testcrossed to A-line testers and new A-lines are testcrossed to R-line testers. New A- and R-lines with good combining ability are used as parents in A × R crosses and the hybrids extensively evaluated in multiple location trials. Evaluation includes the effectiveness of the parents in hybrid seed production. New hybrid cultivars are produced and marketed by private companies.

## Rice

Hybrid rice dates back to 1973 in China when an abortive pollen plant was identified in the wild species *O. sativa f. spontanea* (WA type cytoplasm) on the southern Chinese island of Hainan (Virmani and Ilyas-Ahmed 2007; Guimarães 2012). Initially it was produced by the three-line system just described for sorghum. The first set of male sterile lines was produced in 1973 and the first hybrid was released in 1976. The hybrids outyielded conventional inbred line cultivars by 20 % on average. By 1999, about half (15.5 million ha) of the total rice area in China was planted with hybrids. Hybrids are also grown in India, Philippines, Vietnam, Bangladesh and Indonesia; on average the yield gains over conventional cultivars has ranged from 20 % in Philippines to 30 % in Vietnam. The WA source of sterility-inducing cytoplasm is the most commonly used one in China and elsewhere. The frequency of restorers is around 25 % in *indica* rice germplasm but maintainer frequency is less, 1–15 %, and restorers are almost absent in *japonica* and basmati rice. More recently two-line hybrids have been produced in China where the female parent has environment-sensitive genetic male sterility (EMGS). Such a seed production system is potentially simpler, less expensive and more efficient than the three-line system. Under tropical conditions, where day length differences are marginal, a temperature-sensitive genic male sterile (TGMS) system is considered more useful than a photoperiod-sensitive (PGMS) one. Several TGMS lines have been developed in China, IRRI and elsewhere. For example, TS6 is a spontaneous mutant of RP 2161-106-1-1 identified at the Paddy Breeding Station (PBS), Tamil Nadu Agricultural University (TNAU), India (Latha et al. 2004). As the mutant is recessive, any fertile line can be used as a male parent to develop commercial rice hybrids. Single gene control also facilitates its transfer from one genotype to another. The critical temperature for transformation from

fertility to sterility is above about 26 °C. Stable sterile and fertile phases at different locations were found through sequential sowing so that the proper seasons and locations for the propagation of sterile lines and hybrid seed production could be recommended.

In contrast with rice, there has not been much hybrid wheat development and production worldwide, although some has been done primarily by private companies. They have used either three-line (A, B and R) cytoplasmic male sterility or chemical hybridizing agents (Mergoum et al. 2009a). The latter involves spraying chemicals on the female parents before anthesis to make them male sterile. Later the male-sterile plants are pollinated by wind-borne pollen from untreated male parents. Seeds produced on the female parents are harvested and marketed as hybrid seed.

## ***Tomato***

The tomato was probably domesticated in Mexico, from where it was introduced worldwide from the middle of the sixteenth century onward. In 2012 world production was 162 million tonnes (FAOSTAT data 2012) with China (50.0 million tonnes), India (17.5), USA (13.2), Turkey (11.4), Egypt (8.6), Iran (6.0) and Italy (5.1) the leading producers. The crop is grown as a vegetable in greenhouses and outdoors for fresh consumption. It is also grown outdoors for processing for which determinate growth, a compact habit and uniform ripening are desirable for mechanical harvest. Bai and Lindhout (2007) have provided a useful review of tomato breeding to which the reader is referred for more information. At the end of the nineteenth century, numerous tomato cultivars were available in different colours and for different purposes. They can be regarded as open-pollinated land-races, the products of domestication and early breeding by farmers and growers who extracted seeds from the fruits to produce the next generation. As tomatoes reproduce primarily by self-pollination (94–99 %; Díez and Nuez 2008) these cultivars were true-breeding for most of their characteristics, and did not suffer inbreeding depression. Those produced and maintained by local growers are today regarded as heirloom varieties and are prized for their distinctive flavours, as well as being unique in size, shape and colour. At the beginning of the twentieth century public institutes, mainly in the USA, became more involved in tomato breeding and private companies were formed. Today about a dozen companies are the main global participants in tomato breeding.

During the second half of the twentieth century commercial breeding shifted from open-pollinated cultivars to hybrids, in order to exploit heterosis, provide uniformity and to protect against illegal reproduction. The first hybrid tomato cultivar ‘Single Cross’ was released in 1946. Today nearly all tomato cultivars for the fresh market and an increasing number of cultivars used in processing are hybrids. The parental lines are usually produced from conventional crossing of lines with complementary characteristics followed by inbreeding, with or

without (single seed descent) selection. Only when parental lines become more or less true breeding ( $F_4$  to  $F_6$ ) are crosses made to produce and test hybrids. There has not been the extensive testing for combining ability and use of recurrent selection seen in maize breeding. Specific traits can be incorporated into the parental lines by backcrossing; often from wild relatives. Four phases of breeding goals can be recognized: high yield in the 1970s, shelf-life in the 1980s, taste in the 1990s and nutritional quality in the 2000s. Today many breeding goals focus on characteristics that reduce production costs or ensure reliable production of high yields with high-quality fruits. Hence finding and combining resistances to abiotic and biotic stresses with high yield and quality is important.

Interestingly, hybrid seed production is done by hand crosses, frequently in areas with abundant cheap labour such as East Asia (Horneburg and Myers 2012). Although hand emasculation and pollination is both expensive and time-consuming, it can be justified by the high price of hybrid seed and hence is the most used method. As explained by Díez and Nuez (2008): emasculation of one parent is performed as soon as the petals open, and pollination is performed 2 days later when the female flower becomes fully receptive. The pollen is taken from a fully mature flower of the other parent. Pollen can be stored at room temperature for a few days or for a week or so under refrigerated conditions (9–10 °C). After pollination, covering the flower with a paper bag to protect it from foreign pollen is recommended if the process is done outdoors. Emasculation but not hand pollination can be eliminated by the use of genic male sterility, but has not been used on a large scale despite improvements in its effectiveness (Díez and Nuez 2008). Most genic male sterility is recessive and maintained through backcrossing ( $msms \times Msms$ ). Three types are known: one resulting in sterile pollen ( $ms$  male sterile genes), another in flowers with only a vestigial development of stamens ( $sl$  stamenless genes), and a third group of functionally male sterile mutants; for example the (positional sterile)  $ps-2$  gene which is characterized by fertile pollen but indehiscent anthers.

## ***Brassica napus Rapeseed***

We saw in Chap. 9 that there is normally no effective mechanism to prevent self-pollination in rapeseed; but insect pollination does result in variable amounts of outcrossing, often between 20 and 40 %. Furthermore, *B. napus* is tolerant of inbreeding, and we saw in Chap. 13 that isolated microspore culture from  $F_1$  plants has become an efficient way of producing doubled haploid lines and cultivars. Friedt and Snowdon (2009) explain how this breeding procedure saves 2 years on classical pedigree selection. They also point out that homozygous doubled haploid lines are very suitable for use in hybrid breeding, based on their *per se* performance and combining ability. The extent of heterosis in rapeseed has been analyzed with widely varying results depending on the materials used. Thus Radoev et al. (2008) were able to quote literature reviews reporting averages of around 30 % better

parent heterosis and 30 % mid parent heterosis for seed yield. They in fact found 30 % mid-parent heterosis in their own genetic analysis using a doubled-haploid population from a cross between cultivar "Express" and resynthesized line "R53"; but the hybrid did not outyield cultivar "Express". They concluded that epistasis together with all levels of dominance from partial to overdominance was responsible for the expression of heterosis in rapeseed. However, developments in methods of hybrid production rather than understanding the genetic basis of heterosis have been the driving force behind hybrid rapeseed.

Hybrid cultivars of winter rapeseed and spring canola have rapidly gained in importance since the year 2000 as effective systems were developed for controlled pollination. Yield improvements in European winter rapeseed of up to 15 % have been reported for F<sub>1</sub> hybrids compared to non-hybrids (Friedt and Snowdon 2009). Furthermore, in Germany in 2003/4 the most widely grown winter rape cultivar was the hybrid 'Talent', and by 2007/8 more than 50 % of the 1.5 million hectares of winter rape was planted with hybrid cultivars.

Although numerous CMS systems are available from different sources, their use in oilseed rape breeding is often inhibited by instability, the absence of suitable restorer or maintainer lines, or negative effects of the cytoplasm used to induce the male sterility (Friedt and Snowdon 2009). Restored F<sub>1</sub> hybrids based on the 'INRA-Ogura' CMS system are under increasing production in France, Germany and other European countries. We encountered this system in Chap. 12 on *B. oleracea* hybrids. The 'Ogura' CMS in radish (Ogura 1968) is caused by an aberrant mitochondrial gene, *orf138*, that prevents the production of functional pollen without affecting female fertility. Bannerot et al. (1974) not only introduced the *B. oleracea* nuclear genome into this cytoplasm by crossing and backcrossing, but also the *B. rapa* and *B. napus* nuclear genomes. Subsequently Pelletier et al. (1983) used protoplast fusion to create hybrid cytoplasms for these nuclear genomes in which the chloroplasts of *B. rapa* and *B. napus* were combined with the *R. sativus* mitochondria carrying 'Ogura' male sterility; thus overcoming the problem of cold temperature chlorosis. Male fertility in this system can be restored by a nuclear gene *Rfo* from radish that alters the expression of *orf138* at the post-transcriptional level. This gene was introgressed from radish into rapeseed through intergeneric hybridization followed by extensive backcrossing and pedigree breeding to overcome low female fertility. The result was homoeologous recombination in which an approximately 50 cM region of a *B. napus* was replaced by *R. sativus* (Delourme et al. 1998).

The 'Polima' (*pol*) CMS system resulted from the discovery of male-sterile plants in the cultivar Polima in 1972. It was made workable by screening large numbers of lines in different environments to find stable maintainer genotypes, and has been successfully utilized for hybrid seed production in China. It can be used in both a three-line system (*pol* CMS sterile line, maintainer line and restorer line) and a two-line system (*pol* temperature-sensitive CMS sterile line and restorer line) (Fu et al. 1990). Genetic studies have shown that its fertility restorer gene (*Rfp*) is monogenic dominant and maps to a 29.2-kb genomic region of *Brassica rapa* (Liu et al. 2012). More recently, Ahmad et al. (2014) characterized two cytoplasmic MS

lines (CMS-ARIT121 and CMS-ARIT2) of spontaneous mutant origin in *B. napus*, based on genetic, morphological and molecular studies using a homology-based candidate gene method. They had the same restorer–maintainer relationship as *pol* CMS, but the two were different from each other, with CMS-ARIT2 showing potential as a novel and usable source of CMS.

Finally, hybrid cultivars based on the ‘Male-Sterility Lembke’ (MSL) system are also prominent among winter oilseed rape cultivars in Germany. This is a genic male sterility system based on a spontaneous mutant selected in the nursery of the German breeding company Norddeutsche Pflanzenzucht HG Lembke in the early 1980s. All *B. napus* lines restore fertility, but the details have not been published by the commercial company (Friedt and Snowdon 2009).

Seed production of inbred line cultivars is straightforward, the most important consideration being sufficient isolation from other oilseed rape crops to exclude cross pollination (in Germany, 200 m for basic seed and 100 m for certified seed). In contrast, hybrid production is more complicated. The parental inbred lines have to be propagated under strict isolation. Then F<sub>1</sub> hybrid production is done in alternate strip cultivation of female seed parent and male pollinator lines, usually in the ratio of 3 to 1. Bee hives are often placed in the field to provide pollinators. After the end of flowering, the male parent is physically removed before the hybrid seed is harvested. The field-grown hybrid plants are checked for off-types i.e. plants that are not typical male fertile (restored) plants.

## Sunflower

My final example is sunflower (*Helianthus annuus*) which I have decided to place here rather than under hybrid cultivars, as it follows on from rapeseed. A review of current genetic, genomic and ecological resources in sunflower, including implications for breeding, has been provided by Kane et al. (2013). Here, however, I am going to concentrate on hybrid production, details of which can be found in the review by Fernández-Martínez et al. (2009). One of the reasons for the transformation of sunflower into a major oilseed crop was the commercial production of hybrid seed, based on a cytoplasmic male sterility (CMS) system combined with fertility restoration by nuclear genes. The replacement of open-pollinated cultivars with hybrids is thought to have given a yield increase of about 20 %.

Wild sunflowers have a sporophytic self-incompatibility system that promotes insect-mediated cross-pollination, commonly by bees. The system was maintained in landraces and open-pollinated cultivars, but hybrid breeding has been accompanied by intense selection for high levels of self-compatibility, and most commercial hybrids are virtually self-fertile. Nevertheless, bee pollination is important for the production of hybrid seed. A total of 72 CMS sources had been identified by 2009, most from progenies of crosses between 16 different wild *Helianthus* selections and cultivated lines. Fertility restoration genes have been reported for 34 CMS sources, mainly in wild *Helianthus* species but also in related genera. Detailed inheritance

studies have been conducted for 19 of the CMS sources. Nevertheless, virtually all cultivated sunflower hybrids are currently based on the original source (PET1) found in 1968 by Leclercq (1969) in the progeny of an interspecific cross between *H. petiolaris* and cultivated sunflower. A large number of PET1 fertility restoration lines have been developed, both from wild species and obsolete cultivars. In most cultivated sunflower lines two dominant nuclear genes, *Rf1* and *Rf2*, are necessary to restore male fertility, with *Rf2* present in nearly all sunflower lines (Reddy and Thammiraju 1977). Hybrid sunflower production is therefore a typical three-line system, where one line (A) is the female parent of the hybrid as a result of genetic-cytoplasmic male sterility, the second line (B) is responsible for maintaining the sterility, and the third line (R) is the male parent of the hybrid that restores its fertility through a dominant restorer gene(s). Single cross (A × R) and three-way hybrids [(A × unrelated B) × R] can be produced, with the latter currently preferred when a vigorous female parent is required to reduce the cost of seed production.

The following brief description of hybrid breeding is based on the review by Fernández-Martínez et al. (2009). The source germplasm for producing inbred lines can be older open-pollinated cultivars and populations improved by recurrent selection ( $S_0$  generation), or planned crosses between superior inbred lines ( $F_2$  generation). Furthermore, existing inbred lines can be improved for specific simply inherited traits by backcrossing. Artificial hybrids are produced by emasculation of the female parent followed by pollination with pollen of the desired male parent. Female (B) and restorer (R) populations are generally managed separately to simplify the development of inbred lines and reciprocal full-sib selection has been practised to provide suitable starting material. Inbreeding usually proceeds by self-pollination and pedigree selection for five generations, although bulk selection and single seed descent can be used. Paper or cloth bags are used to isolate the sunflower heads from insect pollination. Greenhouses and off-season nurseries are frequently used to grow more than one generation a year and hence speed up the breeding programme. Testcrosses usually begin at  $S_3$  ( $F_5$ ) for combining ability to an inbred tester being used in commercial hybrids. Conversion of superior B lines to CMS (A) lines by backcrossing also starts at this stage. Finally the best hybrids are sought by making and testing actual A × R hybrids.

Increases of the A, B and R lines are initially accomplished under bags in nurseries to check for purity and stability of the cytoplasmic male sterility and to eliminate off-types. For example, if the identity of individual A- and B-line plants is maintained, then only those B-line plants that produce completely sterile progeny are retained. Further line increases can be done under isolation in screened cages to eliminate bagging. When crosses between the A and B lines are required, a hive of bees is placed inside each cage to eliminate hand crossing. Field increases of the female (A) line involves planting the A and B lines in rows in a ratio 1:1 or 2:1 (A:B). The production field needs to be isolated by at least 6–8 km from commercial fields, and wild populations in countries such as the USA where those are frequent. One hive of bees per hectare is placed in the field. Isolation is also important in the final hybrid seed production to maintain genetically pure hybrid seed. Based on the flight range of honeybees, seed certifying agencies have

established minimum isolation requirements of 1.6–4.8 km between seed production fields and those of commercial sunflower. The ratio of female to male rows usually ranges from 2:1 to 6:1, depending on the pollination ability of the male parent. Staggered sowing of the parental lines may be required to achieve flowering synchronization between the female and R-line. Again 1–4 hives of honeybees per hectare are usually required during the pollinating period. Rouging of off-types before anthesis is essential, as is removal of any male-fertile plants within the male-sterile line. The rows of the male parent need to be removed after pollination and before the female rows are harvested to avoid contamination.

**Part IV**

**Improving High Yielding Cultivars**

# **Chapter 15**

## **Use of Sexual Reproduction in Base Broadening and Introgression**

### **Introduction**

Modern high yielding cultivars have a narrow genetic base in two respects. Firstly, they contain only a portion of the genetic variation present in the landraces from which they were derived. Secondly, cultivated species have only some of the variation present in the wild species from which they were domesticated and often have wild relatives that were not domesticated. Furthermore, study of the evolution of any particular crop is likely to reveal one or more bottlenecks in its history under cultivation. Examples are dispersion of the crop to new environments, adaptation of the crop to different latitudes (daylengths), survival of disease epidemics and new end uses. All of these examples result in the survival of a sample of the original genetic variation, either by chance or by selection for adaptation to new environments and end uses. Despite the superiority of modern high yielding cultivars in many present day farming systems, land races and wild relatives contain genetic variants that are desirable for use in future breeding and which have not been used to date by farmers and plant breeders. This novel genetic diversity will be needed to increase crop productivity and quality, disease and pest resistance and tolerance of adverse or marginal environments. But the modern cultivars have largely replaced the land races and land degradation and changes in use have endangered the habitats of many wild relatives. The Green Revolution of the 1960s and 1970s boosted productivity with new high yielding cultivars of wheat and rice, but contributed to loss of genetic diversity as they replaced the extremely variable land races. In India, for example, farmers historically cultivated around 30,000 rice varieties but now 75 % of their rice comes from just ten varieties (Medina 2012). In the USA, more than 90 % of the fruit trees and vegetables that were being grown at the beginning of twentieth century can no longer be found, with just a few of them maintained in genebanks (Esquinás-Alcázar et al. 2011). Wild relatives once found in forest and savanna have been eliminated by environmental degradation, such as forest clearance for timber or to create more

agricultural land; and also by the re-seeding of pastures with standard commercial seed mixtures and the widespread grazing of cattle, sheep and goats (Hawkes 1991). Hence there is a need to conserve this biodiversity *in situ* in protected areas and by farmers, and *ex situ* in genebanks. There also needs to be an incentive for plant breeders to utilize this biodiversity. As breeders make progress they create an elite germplasm of outstanding genotypes adapted to current farming systems, environments and end uses. These outstanding genotypes are destroyed by crosses with unimproved exotic germplasm. Hence as long as breeders continue to make progress by intercrossing and selection within elite germplasm, they are reluctant to access what appears to be agronomically inferior germplasm. They will do so if absolutely necessary for specific new traits such as disease and pest resistance, and this did in fact happen from early in the twentieth century in the form of introgression breeding. Later mutation breeding, and more recently genetic transformation, has opened up new possibilities. However, by the 1960s there were concerns that in some crops a yield plateau might be reached, or more generally that the rate of progress was slowing down, or that the needs of end users were not being adequately met. Since then there has been interest in a more extensive broadening of the genetic base of crop production, described by some as incorporation to distinguish it from introgression. A review of progress to the end of the twentieth century can be found in the book *Broadening the Genetic Base of Crop Production* edited by Cooper et al. (2001).

## Genetic Resources

As explained in Chap. 1, our modern view of the origin, geography and evolution of cultivated plants dates back to the work done from 1920 to 1940 by Nickolay Ivanovich Vavilov (1940). Today his views on the direct relationship between centres (reservoirs) of diversity and centres of origin of cultivated plants are considered too simplistic. But the collection and conservation of germplasm of cultivated species and their wild relatives in genebanks was a direct outcome of his work, and a legacy which has continued to the present day. I will use potatoes as an example. Wild tuber-bearing *Solanum* species are distributed from the south western United States (38°N) to central Argentina and adjacent Chile (41°S) (Hawkes 1990; Spooner and Hijmans 2001). They are a tremendous resource for potato breeding because of their wide geographical distribution and great range of ecological adaptation. There have been numerous collecting expeditions, from those pioneered by the Russians in the 1920s (Hawkes 1990) to the more recent ones of the 1990s (Spooner and Hijmans 2001). The collecting expeditions led to the establishment of a number of potato germplasm collections worldwide. The world collection has been held at The International Potato Centre (CIP) (<http://www.cipotato.org>) since its creation in Lima, Peru in 1970. CIP has also assembled a collection of 3527 potato cultivars (landraces) native to nine countries in Latin America: Argentina, Bolivia, Chile, Colombia, Ecuador, Guatemala, Mexico, Peru and Venezuela. The other

main potato germplasm collections are the Commonwealth Potato Collection (CPC, Dundee, Scotland), the Dutch-German Potato Collection (CGN, Wageningen, The Netherlands), the Groß Lusewitz Potato Collection (GLKS, IPK, Groß Lusewitz, Germany), the Potato Collection of the Vavilov Institute (VIR, St. Petersburg, Russia), the U.S. Potato Genebank (NRSP-6, Sturgeon Bay, USA) and Potato Collections in Argentina, Bolivia, Chile, Colombia and Peru. Together they comprise the Association for Potato Intergenebank Collaboration and have established an Inter-genebank Potato Database (IPD) ([www.potgenebank.org](http://www.potgenebank.org); or individual genebanks through Google). Germplasm resources for other cultivated plants can be found in *The Handbook of Plant Breeding* (Prohens et al. 2008). They are too numerous to list here, although mention will be made of some resources. My purpose in this chapter is to concentrate on methods of utilization of genetic resources in plant breeding through sexual reproduction. Subsequent chapters will deal with mutation breeding and genetic transformation.

The dangers of a narrow genetic base in modern high yielding cultivars, combined with monoculture on a vast scale, were dramatically illustrated in the US corn-belt when leaf blight disease struck in 1970. Hybrid breeders had relied on a single source of CMS that proved particularly susceptible to the disease. As a result, the US Academy of Sciences produced a major report on the 'Genetic Vulnerability of Major Crops' (National Research Council 1972) which recommended a broadening of the genetic base of major staple crops. Lessons were learnt, and so too were the wider implications of a narrow-genetic base in plant breeding. One consequence was the establishment of The International Board for Plant Genetic Resources (IBPGR, now Biodiversity International) in 1974, which built on the co-ordinated conservation efforts that emerged from the Technical Meeting in Rome in 1961 of the Food and Agriculture Organization of the United Nations (FAO).

Adoption of the Rio (Earth Summit) Convention on Biological Diversity (CBD) in 1992, and the agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS) at the end of Uruguay Round of the General Agreement on Tariffs and Trade (GATT) in 1994, led to an appreciation of the uniqueness of plant genetic resources for agriculture (Esquinas-Alcázar et al. 2011). As a consequence, in 1995 the CBD formally recognized the special nature of agricultural biodiversity, and during the 31st Conference of FAO the International Treaty on Plant Genetic Resources was adopted on 3 November 2001. The Treaty was in fact negotiated within the intergovernmental Commission on Genetic Resources for Food and Agriculture, following the adoption back in 1983 of the International Undertaking on Plant Genetic Resources for Food and Agriculture, within the FAO. In the final treaty, Esquinas-Alcázar et al. (2011) conclude that "neoliberal economic theories prevailed over utopian socialism". Plant breeding always has a political, economic and social context. Indeed, Esquinas-Alcázar et al. (2011) explain that the Treaty covers 64 crops and forages and not the 67 intended because of a conflict over the occupation of China airspace by an aircraft of the United States on 1 April 2001. China withdrew soybean from the Treaty's list and Latin American countries reacted by withdrawing peanut and tomato. China is the primary centre of diversity

of soybeans, the USA and Brazil being leading producers of the crop, whereas peanuts are of great importance in China, and Brazil and Bolivia contain peanut's maximum diversity.

## International Treaty on Plant Genetic Resources

The International Treaty on Plant Genetic Resources for Food and Agriculture came into force on 29 June 2004 and the first session of its Governing Body took place in June 2006 in Madrid. By 1 June 2011, 127 states were contracting parties to the Treaty. The following information is taken from the book edited by Frison et al. (2011), *Plant Genetic Resources and Food Security*, and from the website (<http://www.planttreaty.org/>) to which readers are referred for up to date information. A discussion of the scientific, technical, socio-economic, legal and institutional challenges ahead can be found in the opening chapter of the book (Esquinas-Alcázar et al. 2011). The Treaty aims at recognizing the enormous contribution of farmers to the diversity of crops that feed the world; establishing a global system to provide farmers, plant breeders and scientists with access to plant genetic materials; and ensuring that recipients share benefits they derive from the use of these genetic materials with the countries where they have been originated. The treaty is considered crucial in the fight against hunger and poverty and essential for the achievement of the United Nations' Millennium Development Goals. It recognizes that no country is self-sufficient in plant genetic resources; all depend on genetic diversity in crops from other countries and regions. International cooperation and open exchange of genetic resources are therefore essential for food security; but so is the fair sharing of benefits arising from the use of these resources.

The Treaty's truly innovative solution to access and benefit-sharing is its declaration that 64 of our most important crops, that together account for 80 % of all human consumption, will comprise a pool of genetic resources that are accessible to everyone. On ratifying the Treaty, countries agree to make their genetic diversity and related information about the crops stored in their genebanks available to all through the Multilateral System (MLS) with its Standard Material Transfer Agreement. This cuts down on the costly and time consuming need for breeders to negotiate contracts with individual genebanks. The Multilateral System sets up opportunities for developed countries with technical know-how to use their laboratories to build on what the farmers in developing countries have accomplished in their fields. Under the Treaty and its Multilateral system, the collections of local, national and international genebanks that are in the public domain and under the direct control of Contracting Parties share a set of efficient rules of facilitated access. This includes the vast collections of the Consultative Group for International Agricultural Research (CGIAR), a consortium of 15 international research centres. In addition, the Global Crop Diversity Trust, a complementary funding mechanism to that established by the Treaty, is committed to raise the funds that will endow the genebanks and ensure their continued viability. Those who access

genetic materials through the Multilateral System agree that they will freely share any new developments with others for further research or, if they want to keep the developments to themselves, they agree to pay a percentage of any commercial benefits they derive from their research into a common fund, established in 2008, to support conservation and further development of agriculture in the developing world. The Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization is a supplementary agreement to the Convention on Biological Diversity. It was adopted on 29 October 2010 in Nagoya, Japan, and came into force on 12 October 2014. It provides a legal framework for the fair and equitable sharing of benefits arising out of the utilization of genetic resources. This emphasizes the fact that both the Convention on Biological Diversity and the International Treaty are works in progress.

The International Treaty recognizes the enormous contribution that the local and indigenous communities and farmers of all regions of the world, particularly those in the centres of origin and crop diversity, have made and will continue to make for the conservation and development of plant genetic resources, which constitute the basis of food and agriculture production throughout the world. It gives governments the responsibility for implementing Farmers' Rights, and lists measures that could be taken to protect and promote these rights. These include the protection of traditional knowledge relevant to plant genetic resources for food and agriculture; the right to equitably participate in sharing benefits arising from the utilization of plant genetic resources for food and agriculture; and the right to participate in making decisions, at the national level, on matters related to the conservation and sustainable use of plant genetic resources for food and agriculture. The International Treaty also recognizes the importance of supporting the efforts of farmers and local and indigenous communities in the conservation and sustainable use of plant genetic resources for food and agriculture, including a funding strategy.

### ***Global Crop Diversity Trust***

An essential element of the Funding Strategy of the Treaty is the Global Crop Diversity Trust. It was formally established in October 2004 as an independent organization under international law, a status conferred through the signing of an Establishment Agreement by seven states from five of the regions referred to in the basic texts of FAO. An agreement defining the relationship between the Trust and the Governing Body of the Treaty was signed in June 2006. The Trust's aim is to ensure the long-term conservation and availability of Plant Genetic Resources for Food and Agriculture, with a view to achieving global food security and sustainable agriculture. It has established an endowment fund and receives and distributes other funds. More information can be found in the book chapter by Hawtin and Fowler (2011).

### **Svalbard Global Seed Vault**

The Treaty also led to the Svalbard Global Seed Vault which opened in February 2008 (Hawtin and Fowler 2011). The Vault aims to provide a fail-safe back-up facility for the world's germplasm collections by storing a duplicate set of as much of the world's unique crop genetic material as possible. Its construction was funded by the Government of Norway who own, maintain and administer the facility. It operates through a tripartite agreement between the Norwegian Government, Global Crop Diversity Trust (GCDT) and Nordic Genetic Resource Centre (NordGen), who first identified Svalbard as a suitable location in 1983. The site is a mountain on a remote island in the Svalbard archipelago, halfway between mainland Norway and the North Pole. NordGen manages the vault, overseen by an International Advisory Council. GCDT covers the primary ongoing operational costs. The Vault comprises three chambers over 125 m into the mountainside, each having the capacity to store 1.5 million seed samples. The chambers are artificially cooled to  $-18^{\circ}\text{C}$  with security against breakdown provided by the fact that they are set deep within permafrost at a temperature of  $-3^{\circ}\text{C}$ . Ownership and access rights to the material remain with the depositor. This means that stored seed packages and boxes can be opened and sent only to the original depositor, who is responsible for testing and any subsequent regeneration of material. By March 2010 it housed 522,000 seed samples from 28 institutions in 24 countries (Hawtin and Fowler 2011).

## **Conservation**

Conservation can be divided into *in situ* (on site) and *ex situ* (off site), and the former divided into conservation of landraces by farmers and the conservation of wild relatives in protected areas. The term 'genetic conservation' is often applied to the conservation of the widest range of genetic traits of potential use in plant breeding, particularly with crop wild relatives (Hunter and Heywood 2011).

### **Conservation In Situ**

Protected areas are nature reserves or 'gene parks' designed to cover as much genetic diversity of the species as possible. This is fine where wild relatives possess a restricted distribution, but problematic with a crop like potatoes where its wild relatives are spread over a linear distance of nearly 10,000 miles, although Argentina, Peru, Bolivia and Mexico do harbour 87 % of the wild relatives (Hijmans et al. 2002). Nevertheless, an increasing number of *in situ* conservation projects are underway, as described and explained in the detailed report by Maxted and Kell

(2009) on the *Establishment of a Global Network for the In Situ Conservation of Crop Wild Relatives: Status and Needs* and in the book edited by Hunter and Heywood (2011), *Crop Wild Relatives A Manual of in situ Conservation*. The manual points out that despite a certain momentum in the last 5–10 years, such conservation is still a poorly understood process with only a limited amount of practical experience. Nevertheless, experiences in Armenia, Bolivia, Madagascar, Sri Lanka and Uzbekistan are welcome additions to the literature. While many challenges are technical in nature, others that need to be addressed are political, institutional, cultural, legal and social, including land ownership and sharing the benefits of resources. There is understandably concern about the survival of some crop wild relatives because of climate change. Jarvis et al. (2008) used current and projected future climate data for about 2055, and a climate envelope species distribution model, to predict the impact of climate change on the wild relatives of peanut (*Arachis*, 51 species), potato (*Solanum*, 108 species) and cowpea (*Vigna*, 48 species). Climate change strongly affected all taxa, with an estimated 16–22 % of species predicted to go extinct, most species losing over 50 % of their range size, and the suitable areas for many species becoming highly fragmented. *Arachis* was most affected and *Vigna* least affected. The authors concluded that there is an urgent need to identify and effectively conserve crop wild relatives that are at risk from climate change. While increased habitat conservation will be important to conserve most species, those that are predicted to undergo large reductions in range size should be a priority for collection and inclusion in genebanks.

Landraces can also be conserved *in situ*, with local farmers best placed for the task. For example, CIP has established a 15,000 ha reserve (Cusco Potato Park) high in the Andes of Peru where the indigenous Quechua people grow 1345 landraces of potato, of which 779 were collected locally, 410 were repatriated from CIP and 157 were received through seed exchanges ([www.cipotato.org](http://www.cipotato.org)). Jarvis et al. (2000) produced *A Training Guide for In Situ Conservation On-farm*, to which readers are referred for a thorough discussion of issues and extensive references. Jarvis et al. (2000) point out that on-farm conservation concerns entire agroecosystems, including immediately useful species (such as cultivated crops, forages and agroforestry species), as well as their wild and weedy relatives that may be growing in nearby areas. Within this definition, a wide range of objectives that may shape an on-farm conservation programme are identified and discussed, not just those concerned with landraces. Hawkes (1991) had thought that it would be impossible by *in situ* means to cover the total genetic diversity of wide ranging crops such as the major cereals, legumes and tuber crops. Furthermore, the international scientific community had favoured most effort going into *ex-situ* conservation, but this is changing. Today scientists recognize that each approach has particular advantages and disadvantages in the conservation of crop genetic resources. While *ex situ* conservation is primarily a technical issue of how best to preserve germplasm, the conservation or erosion of genetic diversity in farmers' fields is shaped by a complex range of factors over time. Thus it is relatively easy to identify and access the genetic diversity conserved in a genebank as the material is usually fully documented for the use of plant breeders and other scientists.

However, the removal of genetic material from its natural environment halts the ongoing evolutionary processes which help to make landraces unique and adaptable to changing environments. Likewise for crop wild relatives, Hunter and Heywood (2011) define gene management zones (GMZs) as natural and semi-natural areas set aside for maintaining genetic diversity in a natural setting for the species of interest and associated species. Here they can continue to evolve and generate new variation, some of which may be valuable in breeding programmes in the future.

## ***Conservation Ex Situ***

Most of the problems of *ex situ* storage of seeds in gene banks have been solved for seeds that can be stored for long periods under low humidity and low temperatures (Hawkes 1991). International standards for long-term storage were recognized in 1975, namely minus 18 °C or less (often –20 °C) in hermetic containers at a seed moisture content of 5 ± 1 % wet weight basis (Roberts 1991). There are, however, ‘recalcitrant’ seeds for which the lowering of humidity and temperature results in early death. Many tropical fruits and timber trees fall into this category and require tissue culture or in-vitro storage. Tuber crops or others where clonal propagation is important are often treated in this way with back-up material in field plantings (Hawkes 1991). However, the Commonwealth Potato Collection (CPC), which dates back to 1938, was fully converted to true seed populations in the 1960s. The loss of identity of the native cultivars in the collection was compensated by conserving the alleles present in each accession in a form which is secure against the casual losses suffered by annually planted clones.

The CPC comprises 1500 accessions, two thirds of which are wild species (80) and one third cultivated. The annual rejuvenation cycle involves 60 accessions so that the entire collection is done every 25 years (Fig. 15.1). Sufficient seed to obtain 20 seedlings of each accession is sown in spring in an insect-proof Quarantine Glasshouse, and the plants tested for quarantine organisms (e.g. viruses) that could be present in stored seed. The 20 plants are intercrossed at flowering by applying bulked pollen to all open flowers, repeating the process as necessary. Seed is extracted from ripe berries, dried (6 % moisture), packeted and kept in a dedicated store in conditions (low temperature and humidity) designed to retain viability for many years. A working collection store is maintained at 4 °C and 10 % relative humidity. The CPC database is regularly updated and more information about the collection can be found on the website <http://www.hutton.ac.uk/about/facilities/commonwealth-potato-collection>. Utilization of the collection in potato breeding was reviewed by Bradshaw and Ramsay (2005).

It should be pointed out that rejuvenation methods do depend on the breeding system of the species. Thus out-pollinating vegetables, for example, are isolated from similar taxa in insect-proof compartments to prevent contamination and pollination done by introduced flies. The size of seed sample to be used in regeneration needs to reflect the genetic composition of the accession, with the



**Fig. 15.1** Rejuvenation of part of Commonwealth Potato Collection (*source: SCRI*)

effective population size ( $N_e$ ) chosen to minimize loss of alleles. Twenty plants are used in the CPC, whereas a minimum of 100 have been recommended for perennial allogamous species such as grasses (Johnson et al. 2002). Regeneration is commonly done when seed viability falls below 85 % of the initial viability and the target for storage is often 1500 seeds for self-pollinating species and 3000 for outcrossing ones (FAO 2013).

### ***Genebank Standards and Information***

By 2014, more than seven million samples of seeds, tissues and other plant-propagating materials from food crops, along with their wild relatives, were safeguarded in about 1750 genebanks worldwide (<http://www.fao.org>). In 2013, FAO (2013) produced a new publication, *Genebank Standards for Plant Genetic Resources for Food and Agriculture* (<http://www.fao.org/publications/en/>), aimed at improving conservation of food crops. It outlines voluntary, international standards for genebanks that are designed to guide users in implementing the most appropriate technologies and procedures for the collection, conservation and documentation of crop diversity. The general principles of genebank standards are dealt with under the headings: identity of accessions, maintenance of viability, maintenance of genetic integrity, maintenance of germplasm health, physical security of collections, availability and use of germplasm, availability of information and

proactive management of genebanks. Then the three types of genebank are covered in detail, namely seed banks, field genebanks and *in vitro*/cryopreservation genebanks.

Information on nearly three million accessions can be accessed through Genesys. It is the Internet's largest gateway through which users can discover material in genebanks around the world. Genesys brings together accession-level information on genebank collections from many different sources, including passport, characterization and evaluation data, as well as environmental information from the sites of collection. The Genesys project was started in 2008 and an updated version launched in 2014. The previous 2011 version brought together accession-level data from three major networks:

1. The European Plant Genetic Resources Search Catalogue (EURISCO)
2. CGIAR's System-wide Information Network for Genetic Resources (SINGER)
3. The United States Department of Agriculture's [Genetic Resources Information Network \(GRIN\)](#)

Genesys currently provides access to an estimated one-third of genebank accessions held worldwide. EURISCO, for example, is a search catalogue providing information about *ex situ* plant collections maintained in Europe, based on an European network of *ex situ* National Inventories (NIs). Currently, EURISCO comprises passport data for about 1.1 million samples, representing 5929 genera and 39,630 species from 43 countries.

Before moving on to consider utilization of genetic resources, two topics are worthy of further treatment, the size of sample when collecting germplasm and core collections.

### ***Size of Sample When Collecting Germplasm***

A particularly encouraging theoretical result was obtained by Lawrence et al. (1995) for the sample size when collecting germplasm in the first place. They found that a random sample of 172 plants was sufficient to conserve at a very high probability, virtually all of the polymorphic genes (alleles) provided that their frequency was not less than 0.05, irrespective of the mating system (self-, cross- or mixed pollination). Some actual figures are of help in appreciating this result. The sample size of 172 was derived as follows. In Chap. 9 we derived the following result for a single locus with two alleles (*A* and *a*) at frequencies *p* and *q* under mixed selfing and random mating. At equilibrium the genotype frequencies ( $P_{AA}$ ,  $P_{Aa}$  and  $P_{aa}$ ) are:

Genotypes	<i>AA</i>	<i>Aa</i>	<i>aa</i>
Frequencies	$P_{AA} = p^2 + fpq$	$P_{Aa} = 2pq(1-f)$	$P_{aa} = q^2 + fpq$

$$f = s/(2-s) \text{ where } s \text{ is the proportion of selfing}$$

The probability ( $P$ ) of retaining both alleles in a sample of size  $n$  is  $1 - (P_{AA})^n - (P_{aa})^n$ .

In other words, one must not pick all  $AA$  or all  $aa$  in the sample.

If  $p = 0.95$  and  $q = 0.05$  and  $P = 0.9999$ , then for  $s = 0$  ( $f = 0$ ),  $n = 86$ ; and for  $s = 1$  ( $f = 1$ ),  $n = 172$  and this is considered the worst case scenario. The key choice is setting the probability at 99.99 %.

With four alleles at frequencies 0.85, 0.05, 0.05 and 0.05, the probability of not losing an allele at one locus in a sample of 172 plants is >99.99 % with random mating and >99.95 % with complete selfing. With two alleles at frequencies 0.90 and 0.10 at each of 1000 loci, the probability of not losing an allele in a sample of 172 plants is >99.99 % with complete selfing, and still >99.97 % with 20,000 loci ( $0.999999986^{20000}$ ). If the alleles frequencies are 0.95 and 0.05 the probability does fall to 5.24 % with 20,000 loci and complete selfing; but just 10 % cross-pollination increases it to 54.18 %. The authors also argued that when samples were taken from a number of populations, the size of the sample drawn from each need be no larger than 172 divided by the number of populations visited. These considerations lead us into the concept of core collections.

## ***Core Collections***

As many genebanks became very large, it was suggested in the 1980s that they could be divided into an important core collection, used for evaluation and other purposes, and the rest set aside as a reserve collection (Hawkes 1991). Brown (1989) made the case for core collections but emphasized that they should be seen as an efficient way into the total collection for its greater use. He considered the steps required in setting up a core collection and how population genetics theory could provide a guide to size and hierarchical cluster analysis a guide to content. He argued that conceptually there are four classes of alleles among the total number of accessions, namely the four combinations of common versus rare and widespread versus localized. An allele is common if it has a frequency of greater than say 0.10 in an accession and its common occurrences can be widespread or localized in very few accessions, at the extreme in just one. The same applies to rare alleles with a frequency of less than say 0.10. Common and widespread alleles will almost certainly be included in a subsample of the collection whereas rare and localized ones will inevitably be excluded. That leaves two classes for consideration in formulating sampling strategies. The common but localized ones are likely to be of particular interest to plant breeders because they are likely to include disease and pest resistance and tolerance to abiotic stresses. The genebank curator will want to try to capture these alleles through coverage of the major discontinuities in geographic areas, ecological zones and morphological variation. Lastly, application of the infinite neutral allele model of population genetics, and its sampling theory, to the rare but widespread class leads to a rational target size for a core collection. The

model assumes an infinite number of selectively neutral alleles at a locus where, in the long run, a balance is struck between the mutational input of new alleles and the loss of old alleles by random genetic drift. Under this model, the mean number of different alleles present reaches equilibrium.

The theory can be found in books on population genetics (Gale 1990; Hartl and Clark 1997). The results presented by Brown (1989) suggest that a core collection should be about 10 % of the total collection, but also at least greater than 3000 individuals (e.g. 60 accessions each of 50 propagules). These conclusions are based on the following reasoning. A sample of 3000 individuals is expected to contain 9.2 neutral alleles at a single locus if  $\theta = 1$ , where  $\theta = 4N_e u$ , and  $N_e$  is the effective population size and  $u$  is the mutation rate. This number of alleles is as many distinct alleles at that locus as exist in the total collection with frequency greater than  $10^{-4}$ . Next one plots the lower 95 % confidence limit for the expected number of alleles retained in the core collection as a fraction of the total number present in the total collection, against the percentage of the total collection kept in the core collection. This is done for a single locus, 100 loci and an infinite number of loci. The curves have two distinct phases, a slowly increasing loss of alleles as the percentage retention of the total collection decreases until somewhere below 10 % there is a marked increase in the rate of loss. Hence the rule of thumb is that the size of the core collection should be about 10 % of the entire collection. Allowing accessions to differ in their level of diversity leads to lower levels of variation in the core collection, but the effect is not that marked. Having decided the size of the core collection, the curator needs to choose which accessions to include based on a hierarchical cluster analysis and consideration of the genetic structure of the germplasm collection. The genetic structure will reflect geographic patterns of population differentiation and genomic associations due to linkage disequilibrium, as also discussed by Brown (1989).

### ***Cultivated Potatoes in Latin America: A Core Collection***

Following its creation in Lima, Peru in 1970, The International Potato Centre (CIP) (<http://www.cipotato.org>) assembled a collection of more than 15,000 accessions of potato cultivars (landraces) native to nine countries in Latin America: Argentina, Bolivia, Chile, Colombia, Ecuador, Guatemala, Mexico, Peru and Venezuela. Subsequently duplicate accessions were identified and the number of individual cultivars was reduced to 3527 of which 552 were diploids, 128 triploids, 2836 tetraploids (2644 Group Andigena, 144 Group Tuberosum and 48 hybrids) and 11 pentaploids (Huaman et al. 1997). By 1997 researchers at CIP had already conducted 46,124 evaluations on the collection for the reactions of cultivars to abiotic and biotic stresses and for other desirable traits. A core set of 306 Group Andigena accessions from the 2379 still held at CIP was then established to aid utilization (Huaman et al. 2000a). They were chosen to represent the widest morphological diversity and to maximize geographical representation, following

a cluster analysis of morphological data. Evaluation data were taken into account in choosing the representative accession of each cluster. Isozyme analysis of the entire collection confirmed that the core collection had captured a representative sample of the alleles at nine allozyme loci, with only the loss of alleles whose frequency was less than 0.05 % (Huaman et al. 2000b). A simulation study revealed, however, that a core collection size of 600 would be required to adequately represent allele frequencies and locus heterozygosity (Chandra et al. 2002). Nevertheless, the CIP core collection should be valuable for future breeding both in South America and worldwide. Other genebanks have collections of cultivated species, but the one at CIP is recognized as the world collection. The collection is maintained by clonal propagation in the field and *in-vitro*. In the future, cryopreservation may become an important component of maintaining clonal germplasm collections, thus allowing a reduction in the labour required for routine subculture of *in-vitro* stocks. Much research has been done in recent years with some encouraging results that have been reviewed by Kaczmarczyk et al. (2011).

### **Diversity Fixed Foundation Sets**

The concept of a core collection has been developed further by the generation of homozygous lines from the starting, or founder lines of the core collections. Accessions of outbreeding crops are usually genetically variable populations with heterozygous individuals. In contrast, homozygous lines are genetically uniform and true breeding, both advantages for genetical research and use in breeding. The collections of homozygous lines are called Diversity Fixed Foundation Sets (DFFS). A good example is the *Brassica oleracea* DFFS which has been developed at the Warwick Genetic Resources Unit (WGRU), a genebank that holds the UK collections of vegetable germplasm including brassicas, alliums, lettuce, carrot, radish and celery (<http://www.warwick.ac.uk/go/gru>). The *Brassica oleracea* DFFS was developed from 376 founder accessions of a core collection mostly sourced from genebanks, but also included fixed founder lines used as parents in a number of mapping populations. Plants were grown to flowering, subjected to microspore culture, and doubled haploids produced. Details can be found in the references given in the review by Walley et al. (2012).

## **Utilization of Genetic Resources**

Collection and storage of genetic diversity is not an end in itself, but rather the prerequisite for evaluation and use in plant breeding. Methods of evaluation can be found in the relevant literature. Now it is time to consider utilization: the introgression of genes from crop wild relatives into crops and base broadening through the use of landraces.

## Past Utilization of Genetic Resources

It is not possible to give a comprehensive review of past utilization of crop wild relatives by plant breeders, which dates from around 1900. Furthermore, Maxted and Kell (2009) point out that the earliest farmers probably used natural introgression to improve their crops. For example, subsistence farmers in Mexico annually grew cultivated maize near its wild relatives to facilitate introgression from the wild species into the crop (Hoyt 1988). For most crop species in *The Handbook of Plant Breeding* (Prohens et al. 2008) there are sections on Genetic Resources and Major Breeding Achievements which provide useful information on utilization. Furthermore, Volume 5 on *Fodder Crops and Amenity Grasses* has a whole chapter on Genetic Resources (Boller and Greene 2010). Wild forms of common forage species still exist and so do feral forms that originated from forage crops but escaped to persist in the natural environment. Thus four categories of plant genetic resources are recognised: wild relatives, wild and semi-natural forms of cultivated species, landraces maintained by farmers through repeated cycles of seed harvest and re-seeding and cultivars. The chapter provides an interesting comparison of *in situ* and *ex situ* conservation as well as strategies for use in breeding. Here in this section mention will be made of two general reviews and one crop specific review that provide indicators of the extent of utilization of genetic resources.

Hajjar and Hodgkin (2007) reviewed the use of crop wild relatives in released cultivars of 19 crops of international importance in the 20 years up to 2007. They found that the total number of contributed traits was 55 in tomato, 12 in each of potato and rice, 9 in wheat and 7 in sunflower; with 1–3 traits in each of cassava, millet, maize, lettuce, banana, barley, groundnut and chickpea; but none in soybean, pigeon pea, sorghum, lentil, Phaseolus beans and cowpea. Although there was a strong emphasis on disease and pest resistance, a wider range of characteristics had been introduced than in the past, including drought and salt tolerance, improved quality and cytoplasmic male sterility. Over 60 wild species were identified as having been used. The reader is referred to the review for more details and original references. The authors concluded that crop wild relatives are continually gaining in importance and prevalence, but their contributions to the development of new cultivars remain less than might have been expected given the following factors. There are improved procedures for intercrossing species from different gene pools, advances in molecular methods for managing backcrossing programmes, increased numbers of wild species accessions in gene banks, and a substantial literature on beneficial traits associated with wild relatives. One remaining constraint is the fact that crosses with wild relatives usually produce lines that have poor agronomic performance, and it is not always possible to remove undesirable traits in a backcrossing programme of normal duration and size.

Maxted and Kell (2009) tabulated examples of the use of crop wild relatives in crop improvement programmes for 29 major crops: peanut (*Arachis hypogaea*), oat (*Avena sativa*), beets (*Beta vulgaris*), oil-seed rape (*Brassica napus*), pigeon pea (*Cajanus cajan*), pepper (*Capsicum annuum*), chickpea (*Cicer arietinum*), coffee

(*Coffea arabica*), soyabean (*Glycine max*), cotton (*Gossypium barbadense*), sunflower (*Helianthus annuus*), barley (*Hordeum vulgare*), sweet potato (*Ipomoea batatas*), lettuce (*Lactuca sativa*), tomato (*Lycopersicon esculentum*), cassava (*Manihot esculenta*), apple (*Malus domestica*), tobacco (*Nicotiana tabacum*), banana and plantain (*Musa acuminata* and *Musa balbisiana*), rice (*Oryza sativa*), finger millet (*Pennisetum glaucum*), common bean (*Phaseolus vulgaris*), pea (*Pisum sativum*), sugar cane (*Saccharum officinarum*), potato (*Solanum tuberosum*), sorghum (*Sorghum bicolor*), wheat (*Triticum aestivum*), grape vine (*Vitis vinifera*) and maize (*Zea mays*). They noted the following: for the 29 crop species there were 234 references that reported the identification of useful traits in 183 crop wild species taxa; the degree to which breeders used wild species varied between crops, being particularly prominent in barley, cassava, potato, rice, tomato and wheat; use had not been systematic or comprehensive so most diversity remains to be utilized; the number of publications on use in breeding has gradually increased with 38 % since 1999; traits cited were disease resistance (39 %), pest and disease resistance (17 %), abiotic stress (13 %), yield increase (10 %), cytoplasmic male sterility and fertility restorers (4 %), quality improvement (11 %) and husbandry improvement (6 %); and since the year 2000 the number of attempts to improve quality, husbandry and end-product commodities has increased substantially. Maxted and Kell (2009) also presented case studies for 14 crops in terms of crop importance, wild relatives and their distribution and assignment to gene pools (see next section), known uses of wild relatives in crop improvement, and recommended conservation actions. The 14 crops were: finger millet (*Eleusine coracana*), barley (*Hordeum vulgare*), sweet potato (*Ipomoea batatas*), cassava (*Manihot esculenta*), banana/plantain (*Musa acuminata*), rice (*Oryza sativa*), pearl millet (*Pennisetum glaucum*), garden pea (*Pisum sativum*), potato (*Solanum tuberosum*), sorghum (*Sorghum bicolor*), wheat (*Triticum aestivum*), faba bean (*Vicia faba*), cowpea (*Vigna unguiculata*) and maize (*Zea mays*).

Schneider et al. (2008) reviewed the results of wheat—*Aegilops* hybridization and gene transfer from wild *Aegilops* species to cultivated wheat (*Triticum* spp.). The *Aegilops* genus consists of 11 diploid, 10 tetraploid and 2 hexaploid species. They are found in Mediterranean climates and have great adaptability, being indigenous from the Canary Islands to the western part of Asia, in Afghanistan and West China. *Aegilops* and *Triticum* cross naturally as they are closely related genera and a large number of artificial hybrids were developed by breeders during the twentieth century. After backcrossing to cultivated wheat, chromosome addition and substitution lines were developed as precursors of translocation lines for the stable incorporation of single segments of *Aegilops* chromosomes and hence agronomically important traits. We will look at translocation breeding in more detail later in this chapter. A large number of genes have been transferred from *Aegilops* species to cultivated wheat, including those for resistance to leaf rust, stem rust, yellow rust and powdery mildew, and various pests such as cereal cyst nematode, root knot nematode, Hessian fly and greenbug. The review contains extensive references to all of this work, as well as an assessment of some difficulties that were encountered; for example, gametocidal genes in some *Aegilops* species.

## Gene Pools

When considering genetic resources, plant breeders are interested in the crossability of cultivated plants with their wild relatives and trying to overcome barriers to hybridization. The ‘Gene Pool System’ was developed by Harlan and de Wet (1971) to help plant breeders think about varying degrees of inter-fertility between related taxa, and is often referred to in books on crop wild relatives (Hunter and Heywood 2011). They recognized primary, secondary and tertiary gene pools of crops as follows.

*Primary*: hybridization with crop species easy; hybrids generally fertile (corresponds to traditional concept of biological species, with subspecies of cultivated races and wild races of conspecific (same species) wild progenitor, e.g. *Beta vulgaris* ssp. *vulgaris* (cultivated beets) and its progenitor *Beta vulgaris* ssp. *maritima* (wild sea beet); *Zea mays* (the cultivated races of maize) and its progenitor *Zea mays* ssp. *parviglumis* (wild annual teosinte).

*Secondary*: hybridization possible but difficult; hybrids weak with low fertility but utilization can be achieved by conventional plant breeding methods (includes non-conspecific wild relatives as well as other cultivated species in same genus).

*Tertiary*: hybridization difficult; hybrids lethal or completely sterile so utilization requires special techniques such as bridging crosses, embryo rescue, induced polyploidy and protoplast fusion (includes more distant wild and cultivated relatives, often in different genera).

Harlan and de Wet (1971) considered the primary and secondary gene pools to be the ones breeders normally used, with the tertiary gene pool defining the extreme outer limit of the potential gene pool of a crop. They did, however, recognize that use of the tertiary gene pool could increase if new techniques became available. They pointed out that wheat has over 35 species in its secondary pool (some authors place all *Aegilops* species and *Amblyopyrum muticum* in secondary pool) and a substantial tertiary one; barley has no secondary pool (although some authors place *Hordeum bulbosum* in secondary pool, with *Hordeum vulgare* ssp. *spontaneum* in the primary pool) and a rather small tertiary one; and soybean lacks both a secondary and tertiary pool. They thought that in general different ploidy levels should be regarded as different gene pools and that the primary pool could be usefully divided into races and subraces. But they did also recognize that classification for each crop needed to be considered on its own merits. More details can be found in the book by Harlan (1992) *Crops and Man*, Second Edition, where Harlan now recognizes an entry to the tertiary genepool of soybean. Spillane and Gepts (2001) have argued that a fourth gene pool is now required to reflect recent technological advances in plant molecular biology and genomics. They also review factors that are barriers to deliberately promoting useful gene flow from the secondary and tertiary genepools into the primary cultivated genepool.

*Quaternary*: Any useful genes which can be harnessed from species sexually incompatible with the crop species using genetic transformation.

This fourth gene pool recognizes the unity and evolutionary history of life on earth. As mentioned in Chap. 5, the genetic code is virtually universal. Barriers to gene flow between organisms and reproductive isolation are the products of evolution and an ongoing process.

## Protoplast Fusion in *Nicotiana*

The potential use of protoplast fusion to overcome barriers to sexual hybridization was first demonstrated in *Nicotiana* in 1972 by Carlson et al. (1972). Protoplasts of *N. glauca* and *N. langsdorffii* were isolated from leaf mesophyll cells, fused, and induced to regenerate into plants. The somatic hybrids were recovered from a mixed population of parental and fused protoplasts by regeneration on a medium on which only hybrid protoplasts could regenerate into calli. The recovered calli formed rudimentary shoots and leaves in culture, but failed to form roots. In order to obtain further differentiation of presumed hybrid tissue, the regenerated shoots were grafted onto the freshly cut stem surface of young plants of *N. glauca*. The somatic hybrids produced flowers and fertile seed capsules. The biochemical and morphological characteristics of three hybrids were analyzed in detail and found to be identical to those of the sexually produced amphiploid. Furthermore, the hybrids had a somatic chromosome number of 42, which was the summation of the diploid somatic numbers of the parental species (24 + 18). Since then, several somatic hybrid and cybrid plants have been reported in the genus *Nicotiana*. It therefore seems appropriate to choose a recent one as an example of the value of protoplast fusion in plant breeding (see Liu et al. 2005 for other examples).

Patel et al. (2011) produced somatic hybrid plants of *Nicotiana × sanderae* (+) *N. debneyi* with fungal resistance to *Peronospora tabacina*. *Nicotiana × sanderae* is a sexual hybrid between *N. alata* and *N. forgetiana* (both  $2n = 2x = 18$ ) which is cultivated for its attractive flowers that range from white, through green to pink and red, lines of uniform colour being selected by breeders for commercial production. The hybrid and its parental species are susceptible to blue mould, caused by *Peronospora tabacina*, a fungal disease that devastates ornamental tobaccos in bedding schemes and under glass. *N. debneyi* is a wild Australian species indigenous to Rockingham Bay in Queensland that was identified as a potential source of resistance to *P. tabacina*. It is an allotetraploid ( $2n = 4x = 48$ , XXYY genome designation) that has two genomes distinct from those of tobacco, *N. tabacum* ( $2n = 4x = 48$ , SSST). Although *N. debneyi* can be crossed with *N. tabacum*, the species is sexually incompatible with *N. × sanderae* making it impossible to transfer disease resistance into the ornamental *N. × sanderae* by conventional breeding. The somatic hybrids were produced by electrofusion of mesophyll protoplasts. Eighty-nine plants were regenerated from protoplast derived cell-colonies and transferred to a glasshouse where they flowered. Details of the techniques involved can be found in the paper. Five plants were confirmed as somatic hybrids by their intermediate morphology compared with parental plants, whereas the remaining

84 plants resembled *N. debneyi* and were discarded. The hybrids had chromosome complements of 60 or 62 which were identified to parental genomes by genomic *in situ* hybridization (GISH). All 18 chromosomes from *N. × sanderae* were present along with 42 or 44 from *N. debneyi*. Four or six chromosomes of one ancestral genome of *N. debneyi* had been eliminated during culture of electrofusion-treated protoplasts and plant regeneration. Both chloroplasts and mitochondria of the somatic hybrid plants were probably derived from *N. debneyi*. As all somatic hybrid plants were fertile and stable over three selfed generations, with seed progeny completely resistant to infection by *P. tabacina*, they should be valuable parents for use in breeding programmes in ornamental tobacco.

## Taxonomic Classification and Crossability in the Potato

### *Taxonomy*

The cultivated potato and its wild relatives provide a good example of crossability in relation to taxonomic classification. The taxonomy of wild tuber-bearing *Solanum* species is complicated and under continuous revision. Hawkes (1990) recognized 219 wild tuber-bearing species and arranged them into 19 series of subsection *Potatoe* of section *Petota* of subgenus *Potatoe* of genus *Solanum*. Hawkes also recognized a further nine closely related non-tuber-bearing species that he grouped into two series of subsection *Estolonifera*, but these have been excluded from section *Petota* in more recent taxonomic reviews, leaving a section comprising all tuber-bearing species. In contrast, the latest summary by Spooner and Salas (2006) recognizes 188 wild potato species for section *Petota* that are grouped into four clades, based on plastid DNA, rather than 19 series. The number of species may be further reduced in the future, and clade composition based on chloroplast DNA may change as extensive nuclear DNA sequence data become available. The wild species form a polyploid series from diploid ( $2n=2x=24$ ) to hexaploid ( $2n=6x=72$ ) in which only diploid cytotypes have been found in 123 species and only polyploids in 43 species (Hijmans et al. 2007). Spooner et al. (2004) summarised the putative genome compositions (A, B, C, D and P) of the polyploid species, but it was clear that further research was required to resolve their origins. Nearly all of the diploid species are outbreeders, with a single S-locus, multiallelic, gametophytic self-incompatibility system (Dodds 1965), whereas the tetraploids and hexaploids are mostly self-compatible and were thought to be allopolyploids based on bivalent pairing (Hawkes 1990). However, GISH analysis by Pendinen et al. (2012) revealed both auto- and allopolyploid origins of the hexaploids. Thus *S. hougasii* is an allopolyploid with one AA component genome, another BB component genome and a third genome more closely related to *P. S. demissum* and *S. albicans*, in contrast, are autoploids with all three chromosome sets related to the basic A genome. The GISH results also revealed that *S. acaule* is an

autotetraploid with the basic A genome, whereas *S. stoloniferum* is an allotetraploid with A and B genomes (Pendinen et al. 2008). But where does all of this leave us when it comes to crossability, the real concern of potato breeders.

## Crossability

The crossability of potato species has been determined through artificial pollinations done over many years. The results can be explained primarily but not exclusively in terms of Endosperm Balance Number (EBN), which can be regarded as the effective rather than the actual ploidy of the species (Johnston et al. 1980). In crosses between species with the same EBN, the hybrids have a normal endosperm for nourishing the hybrid embryo whereas in crosses between species with different EBNs the endosperm degenerates. The endosperm develops following the fusion of a sperm nucleus from the male parent with two polar nuclei from the female parent to give a triple fusion nucleus. It is the genetic composition of this nucleus that is important. Under the EBN hypothesis the endosperm is normal when the three nuclei have the same EBN and hence a 2 maternal to 1 paternal ratio of endosperm balance factors. Attempts have been made to understand the genetic and biological basis of the EBN concept, but EBNs are determined experimentally relative to species assigned an arbitrary value of 1. The main groups of species are as follows (Hawkes and Jackson 1992).

22 species:	Diploid	EBN = 1	e.g. <i>S. bulbocastanum</i> and <i>S. commersonii</i>
178 species:	Diploid	EBN = 2	e.g. <i>S. chacoense</i> , <i>S. microdontum</i> , <i>S. spegazzinii</i> and <i>S. vernei</i>
11 species:	Tetraploid	EBN = 2	e.g. <i>S. acaule</i> and <i>S. stoloniferum</i>
<i>S. tuberosum</i> :	Tetraploid	EBN = 4	
8 species:	Hexaploid	EBN = 4	e.g. <i>S. albicans</i> , <i>S. demissum</i> and <i>S. hougasii</i>

Taxonomic species within these crossability groups have evolved by means of geographical and ecological isolation rather than by genetic incompatibility and hence hybridizations are usually successful.

Today potato breeders can usually achieve sexual hybridization between *S. tuberosum* and its wild relatives by manipulation of ploidy with due regard to EBN (Ortiz 1998, 2001; Carputo and Barone 2005; Jansky 2006). EBN can be doubled meiotically through the unreduced  $2n$  gametes produced from several naturally occurring recessive meiotic mutations that are common in *Solanum* species (Tai 1994). It can also be doubled mitotically through the use of colchicine for chromosome doubling, something which can also occur naturally during callus culture. EBN can be halved by haploidization using *in vitro* androgenesis (anther culture) or, more commonly in *S. tuberosum*, by parthenogenesis using pollinations with particular clones of diploid *S. phureja* (Wenzel 1994; Veilleux 2005). The first such haploids of tetraploid *S. tuberosum* were produced in 1958 (Hougas

et al. 1958) and are often called dihaploids. Embryo rescue can also be used to secure a hybrid where embryo abortion is due to a defective endosperm (Jansky 2006).

As the largest compatibility group is the EBN = 2 one, it is now common for potato breeders to secure tetraploid hybrids from  $4x$  (*S. tuberosum*)  $\times 2x$  ( $2x$  *S. tuberosum*  $\times$  wild species) crosses in which an unbalanced endosperm prevents the development of triploid embryos. There are, however, other barriers to hybridization, such as interspecific pollen-pistil incompatibility and nuclear-cytoplasmic male sterility, although fertility restorer genes have been found for the latter (Jansky 2009). Unilateral incompatibility is known to occur when a self-incompatible (SI) species is pollinated by a self-compatible (SC) one so that *S. verrucosum* (SC female)  $\times S. phureja$  (SI male) is successful, but the reciprocal cross fails (Hermansen 1994; Jansky 2006). Sometimes incompatible pollen can be helped to achieve fertilization through a second pollination with compatible pollen, a technique known as mentor pollination (Hermansen 1994; Jansky 2006). These phenomena have been reviewed by Camadro et al. (2004) in the context of how sympatric species maintain their integrity. From time to time potato breeders have unexpected successes and failures when attempting to overcome barriers to hybridization.

Hybridizations with diploid 1EBN and tetraploid 2EBN can be achieved as follows (Carputo and Barone 2005):

$2x$  (1EBN) wild species

$\downarrow$  (colchicine or *in vitro* doubling)

$4x$  (2EBN) wild species  $\times 2x$  (2EBN) *Tuberosum*  $\times 2x$  (1EBN) wild species

$\overbrace{\quad\quad}$        $\overbrace{\quad\quad}$  ( $2n = 2x$  pollen)

$4x$  (4EBN) *Tuberosum*  $\times 3x$  (2EBN) hybrids  $\times 4x$  (4EBN) *Tuberosum*

$\downarrow$  ( $3x$  pollen)( $2x$  eggs)  $\downarrow$

$5x$  (4EBN) hybrids     $4x+$  (4EBN) hybrids

The  $3x$  (2EBN) hybrids produce unreduced pollen (36 chromosomes) from parallel spindle mutation but egg production involving omission of second meiotic division can produce eggs with 24–48 chromosomes and hence hybrids with 48–72 chromosomes. The 4EBN hybrids will cross with tetraploid *Tuberosum* cultivars. An example is the breeding research by Barone et al. (2001) using *Solanum commersonii*, a source of several valuable traits including resistance to tuber soft rot caused by *Pectobacterium carotovorum* (formerly *Erwinia carotovora*), high dry matter content of tubers and tolerance of freezing. They produced triploid hybrids between an accession of *S. commersonii* ( $2n = 2x = 24$ , 1EBN) doubled ( $2n = 4x = 48$ , 2EBN) by *in vitro* tissue culture of leaf discs, and diploid

Phureja-Tuberosum hybrids ( $2n = 2x = 24$ , 2EBN). They found that chromosome numbers in the offspring of their  $3x$  (2EBN) hybrids  $\times 4x$  (4EBN) Tuberosum crosses ranged from 58 to 67. They chose a 60-chromosome hybrid for the next backcross to  $4x$  (4EBN) Tuberosum which resulted in progeny with 48–57 chromosomes. Finally they chose a 48-chromosome hybrid for their third backcross to  $4x$  (4EBN) Tuberosum and secured 48-chromosome offspring.

Potato breeders can also use somatic (protoplast) fusion to achieve difficult or impossible sexual hybridizations, and in ploidy manipulation to achieve maximum heterozygosity (Wenzel 1994; Veilleux 2005). Protoplast fusion can be induced chemically by the polycation polyethyleneglycol (PEG) or via electrofusion, which is now the preferred method. Somatic hybrids derived from the same fusion combination do show genotypic and phenotypic variation and hence require screening for the desired product. Somatic fusion has allowed the production of fertile hexaploid hybrids between tetraploid *S. tuberosum* (EBN = 4) and diploid EBN = 1 species, such as *S. bulbocastanum*, which has a major gene for broad spectrum resistance to late blight (Helgeson et al. 1998; Naess et al. 2000).

## Introgression in Potato

The potatoes which were introduced from South America into Europe and from there to the rest of the world, lacked adequate resistances to the diseases and pests which became problems as the crop increased in global importance. Introgression breeding started in 1908 in Germany and 1909 in Great Britain with the recognition of *S. demissum* as a source of genes for resistance to late blight, caused by the oomycete pathogen *Phytophthora infestans* (Muller and Black 1951). By the end of the 1980s just 6 out of the 219 wild species had been used extensively in the breeding of successful cultivars in Europe, together with cultivated *S. tuberosum* Groups Andigena and Phureja (Ross 1986). *S. demissum* and *S. stoloniferum* had provided resistances to late blight and viruses, *S. chacoense* and *S. acaule* resistances to viruses, and *S. vernei* and *S. spegazzinii* resistances to potato cyst nematodes. Likewise, just three species, *S. demissum*, *S. chacoense* and *S. acaule*, had been used extensively in North America (Plaisted and Hoopes 1989). The resistances to viruses and cyst nematodes proved valuable in crop production, whereas the *S. demissum*-derived *R*-genes failed to provide durable resistance to late blight either singly or in combination due to the evolution of new races of *P. infestans* (Malcolmson 1969). Currently there is much debate over whether or not the *R*-genes for late blight resistance being found in other wild species such as *S. bulbocastanum* will be more durable *per se* or can be deployed in a more durable way (Goverse and Struik 2009). Wild potato relatives would therefore appear to be a largely untapped source of potentially desirable genes, but to date it has proved difficult to successfully utilize them in potato breeding.

Introgression is essentially backcrossing, and in the past it took from three to seven backcrosses to transfer a major dominant resistance gene into a successful cultivar (Ross 1986). Thus cultivar Pentland Ace (the R3 differential, with gene *R3a*) was released from SPBS (Scottish Plant Breeding Station) in 1951 after just three backcrosses to *S. tuberosum*, starting with clone 735 [*S. rybinii* = *S. phureja* (2x) × *S. demissum* (6x)] in 1937. Clone 735 was one of a few artificial tetraploid seedlings secured by hybridization despite the species having different endosperm balance numbers (Bradshaw 2009a). The commercially more successful cultivar Pentland Dell (with genes *R1*, *R2* and *R3*) traces its pedigree back to clones 429a8 [*S. demissum* (6x) × cv. The Alness (4x)] and clone 735 and required five generations. It entered ware production in Great Britain in 1963 when race 4 was the prevalent race of *P. infestans* but succumbed to late blight in 1967. Nevertheless, it went on to become one of SPBS's most successful cultivars and is still grown in Great Britain for making French fries, albeit under the protection of fungicides.

The recurrent parent(s) will be from tetraploid Group Tuberous when the starting material is one of the following: somatic hybrids between tetraploid *S. tuberosum* and primitive diploid (1EBN) species such as *S. bulbocastanum*; autotetraploids produced by colchicine treatment of diploid (2EBN) species such as *S. vernei*, as done at the Scottish Plant Breeding Station in 1957 (Bradshaw 2009a); and hexaploid (4EBN) species such as *S. demissum*. In contrast, the recurrent parent(s) will be from diploid (dihaploid) Group Tuberous, diploid Group Phureja or diploid Group Stenotomum when introgression is from a 2EBN wild species. A comparison of tetraploid and diploid introgression was made by Bradshaw and Ramsay (2005). The complete elimination of unwanted alleles and whole chromosomes from the wild species is faster at the diploid level, unlike the change in allele frequency which is the same. Furthermore, where introgression is performed at the tetraploid level, the result may not be a genotype with 48 Tuberous chromosomes of which one or more has the introgressed gene(s). This need not affect the performance of the genotype and its vegetative propagation but could affect its fertility and use as a parent for further breeding. The result of an introgression from *S. brevidens* was a high yielding clone, C75-5+297, with resistances to both soft rot of tuber and early blight (Tek et al. 2004). Molecular and cytogenetic analyses revealed that C75-5+297 had 47 chromosomes, including four copies of chromosome 8, three from potato and one from *S. brevidens* that was the only part of the wild species genome present. In contrast, Barone et al. (2001) did verify that their introgression of *S. commersonii* into *S. tuberosum* had resulted in a genotype with 48 chromosomes. Furthermore, with species specific molecular markers, they confirmed the occurrence of recombination between homoeologous chromosomes, the extent of the wild genome carried in each backcross, and the efficiency of introgressing useful genes. By the third backcross, the mean value of *commersonii* specific markers was 28.5 % and two out of 20 genotypes combined resistance to tuber soft-rot with other desirable characteristics.

The hybrids between haploids of *S. tuberosum* and diploid wild species with an EBN of 2 will often form tubers in long-day growing conditions, and hence can be evaluated in long days (Jansky et al. 2004; Jansky 2009). When they also produce  $2n$  gametes by FDR (first division restitution), much of the genetic diversity of the wild species can be efficiently transferred to the tetraploid offspring from  $4x \times 2x$  crosses and results in about 25 % of the wild species genes in the final product (Tai 1994; Jansky 2009). However, sometimes such haploid-wild species hybrids need to be improved before they are used in  $4x \times 2x$  crosses, for example, by population improvement by recurrent selection. In other words, we are moving away from introgression of a few genes towards base broadening or incorporation as defined later in this chapter.

As potatoes are heterozygous outbreeders, use of the same recurrent parent during introgression results in a self of the recurrent parent and hence inbreeding depression. This can be avoided by using different *Tuberosum* parents for each backcross, but results in an entirely new cultivar, which may or may not be the desired outcome. The only way to introduce a gene into a known cultivar is by genetic transformation, which is discussed in the chapter after next. It also has the advantage of no linkage drag and hence should be cleaner and faster.

## Introgression in Sugar Beet

Sugar beet (*Beta vulgaris* ssp. *vulgaris*) provides another early example of introgression breeding. It is a relatively new agricultural crop dating from the early 1800s (Biancardi et al. 2010). Initially production was in the temperate climate of Northern Europe which was relatively disease free. As production moved into warmer areas, endemic diseases were encountered that severely limited yield and for which there were no known resistances. One of the first successful attempts to use exotic germplasm was in the Po Valley of Italy, where the high humidity and warm night temperatures provided optimal conditions for cercospora leaf spot (CLS) caused by the fungus *Cercospora beticola*. In 1913 Ottavio Munerati recognized the potential of the wild sea beet (*Beta vulgaris* ssp. *maritima*) growing in the Po Delta as a source of resistance to CLS. The resulting 'Rovigo' breeding programme produced germplasm and cultivars Cesena and Mezzano that have been adopted worldwide as the source of most CLS-resistant germplasm in use today (Biancardi et al. 2010). Repeated backcrossing was required to reduce the undesirable traits of sea beet, and further selection was required in Italy and the USA; but useful commercial cultivars were the result. The resistance is thought to result from at least four or five alleles with variable effects (recessive effects); in other words it is quantitative resistance. A comprehensive review of introgression from wild relatives of sugar beet has been provided by Panella and Lewellen (2007), including collection, maintenance and evaluation of germplasm.

## Introgression in Onion

Shigyo and Kik (2008) reported that 27,000 *Allium* accessions were held in genebanks worldwide, of which 46.7 % were onion/shallot (*A. cepa*). Although there are over 750 *Allium* species, edible and wild, only a few species can be crossed directly with onion, namely *A. cepa*, *A. vavilovii*, *A. galanthum* and *A. roylei*. Hence these species comprise the primary gene pool of onion. *A. fistulosum* and its progenitor *A. altaicum* have been put in the secondary gene pool as *A. roylei* can act as a bridging species between onion and *A. fistulosum*/ *A. altaicum* (others might argue for tertiary gene pool), and *A. pskemense*, *A. oschaninii* and a further 20 species in the tertiary gene pool. The genebank accessions are dominated by *A. cepa*, with other wild relatives present only in limited numbers.

Downy mildew, caused by *Peronospora destructor*, is a serious fungal disease in onions which causes considerable yield losses every year worldwide (Shigyo and Kik 2008). The disease is controlled chemically (about seven times during the growing season), but it is considered desirable to reduce the amount of fungicides being used and hence their possible environmental impact. In-built resistance is needed, but has not been found within onion germplasm. However, complete resistance to the disease was found in *Allium roylei* in 1990, and has been introgressed into onion, albeit after overcoming some difficulties, as reported by Scholten et al. (2007). A partially fertile interspecific hybrid was secured by hand pollination of cultivar Jumbo with pollen from *A. roylei* Stearn accession PI243009. The hybrid was resistant and one plant was backcrossed with cultivar Hylight to produce a first backcross ( $F_1BC_1$ ) population which displayed a 1:1 segregation of resistance to susceptible plants, suggesting a dominant resistance allele (*Pd1*). Resistant plants were backcrossed with cultivar Maxima to produce the second backcross generation ( $F_1BC_2$ ) which was released to various breeding companies. A RAPD marker closely linked to the resistance allele (2.6 cM) was found through bulked segregant analysis and converted into a SCAR marker (on distal end of chromosome 3) for use by the breeding companies. The breeding companies did three further backcrosses using different onion cultivars to produce fifth backcross populations ( $F_1BC_5$ ). Two or three generations of selfing ( $F_1BC_5S_2$  and  $F_1BC_5S_3$ ) followed to fix the resistance in the homozygous state. Resistant plants were selected in the field and checked for resistance in a greenhouse following inoculation. Commercial onion cultivars are now being marketed after two setbacks, but it is too early to report on the durability of the resistance. The first setback concerned the SCAR marker which became increasingly difficult to use and finally lost its discriminating power between resistant and susceptible plants. Since then four tightly linked AFLP markers have been identified. The second setback was in making the homozygous introgression lines when the initial attempts resulted in lethality due to a recessive lethal factor linked to the resistant allele. Subsequently a further 71 (making total of 215)  $F_1BC_5S_2$  resistant plants were selected and selfed, and one completely resistant progeny found. GISH analysis showed that the

chromosomal region harbouring the resistance locus was located on the distal end of chromosome 3, and was the only remaining piece of *A. roylei* in the nuclear background of onion. The viable and resistant true breeding offspring had a smaller and more distally located introgression fragment. In other words, the linkage drag had been broken. The homozygous lines can be used in the breeding of F<sub>1</sub> hybrid cultivars as the resistance is dominant.

## Introgression in Lettuce and Tomato

Lettuce and tomato are inbreeding species, in contrast to potato and onion. Hence genes can be introgressed into a single cultivar, and it is worth looking at a few examples.

### **Lettuce**

Lettuce (*Lactuca sativa*) is a major fresh vegetable whose leaves are commonly used in salad mixtures and sandwiches and whose stems are also consumed cooked, raw, pickled, dried or as a sauce. Production is dominated by China and the USA. The crop is covered by Mou (2008) in the *Handbook of Plant Breeding Volume 1*. The crop probably originated in the eastern Mediterranean basin (e.g. Egypt) over 4500 years ago. Out of about 100 *Lactuca* species, just four can be crossed to each other by conventional hybridization: *L. sativa*, its direct wild ancestor *L. serriola*, *L. saligna* and *L. virosa*. They are all self-fertilized diploids ( $2n = 2x = 18$ ). *L. sativa* and *L. serriola* are easy to cross and the latter has been used extensively in lettuce breeding. Crosses of these species with *L. saligna* can be completely or partially sterile (so secondary gene pool), but the latter has been used in breeding. Currently there is interest in this species as a source of non-host resistance to the oomycete *Bremia lactucae*, the cause of downy mildew, as major dominant (*Dm*) *R*-genes have failed to provide durable resistance. The breeding strategy has been explained by Jeuken (2012). Many interspecific hybrids were made but F<sub>2</sub> seeds were secured from just 4, of which 126 was largest population size. A molecular marker linkage map was constructed. As plants displayed reduced vigour and fertility, single seed descent was not pursued. Rather, 29 fertile backcross inbred lines were produced, each with a single introgressed chromosomal segment from *L. saligna* in *L. sativa* genetic backgrounds, and representing 96 % of the *L. saligna* genome. The lines were developed by 4–5 backcrosses and 1–2 generations of selfing, with marker assisted selection in later generations to select for the fewest chromosome segments of *L. sativa*. The whole process took 3 years, with 2–3 generations per year in the greenhouse. All lines were extensively evaluated for their resistance to races of *B. lactucae*, including field tests at various locations. A combination of three introgression segments (carrying three recessive QTLs) from

*L. saligna* in a *L. sativa* background gave full resistance to all *B. lactucae* isolates in field tests (Zhang et al. 2009a). Their combined resistance was greater than their individual resistances. However, more breeding and research is required as some chromosomal regions for other detected QTLs were obtained only in a heterozygous state, homozygotes apparently being lethal. We will look at molecular marker-assisted introgression in more detail in the next main section.

Crosses of *L. sativa* with *L. virosa* result in highly sterile hybrids, but a fertile amphidiploid can be produced by treatment of the flowers with colchicine to double the chromosome number. *L. virosa* has also been used in lettuce breeding. Mou (2008) gives the example of quality improvement in crisphead lettuces in California and Arizona. Thompson made a cross between PI 125130 (*L. virosa*) and a line derived from a complex *L. sativa*–*L. serriola* cross, treated the hybrid with colchicine, backcrossed the fertile amphidiploid to cultivated lettuce, and selected diploid cultivar Vanguard which was released in 1958.

## **Tomato**

The tomato is covered by Díez and Nuez (2008) in the *Handbook of Plant Breeding Volume 2*. Originally classified as *Lycopersicon esculentum*, it has been renamed *Solanum lycopersicum*. The tomato was probably domesticated in Mexico and its most likely ancestor is the wild cherry tomato, formerly *Lycopersicon esculentum* var. *cerasiforme*. Domestication saw a change from partial allogamy to strict autogamy, as well as an increase in fruit size, so that the cultivated tomato is a self-fertilized diploid species ( $2n = 2x = 24$ ). The wild relatives of tomato (also  $2n = 2x = 24$ ) are native to western South America and have been used extensively since the 1940s as sources of disease (particularly fungi and viruses) and pest (for example, nematodes) resistance. Díez and Nuez (2008) list 12 wild relatives, their crossability to cultivated tomatoes, and examples of genes introgressed into modern cultivars. The former *Esculentum* group of seven crossable species includes the cultivated tomato and self-compatible *S. pimpinellifolium*, as well as *S. neorikii*, *S. habrochaites* and *S. pennellii*. The former *Peruvianum* group comprises self-incompatible *S. peruvianum* and *S. chilense* for which embryo rescue is usually required to secure crosses of these species with the cultivated tomato. *S. lycopersicoides* is a wild nightshade native to Chile and Peru that possesses many traits of potential interest to tomato breeders. Sexual hybridization with cultivated tomato was achieved in the 1950s but the vast majority of hybrids were simultaneously male-sterile and unilaterally incompatible with *L. esculentum*. Virtually no progress with introgression was made until Chetelat et al. (1997) secured by chance a partially male-fertile hybrid (using embryo culture) between tomato and a newly collected population of *S. lycopersicoides*. This demonstrates that the breeders' gene pool can still be widened in unexpected ways as more research is done.

Bai (2012) describes a recent example of introgression breeding in The Netherlands, prompted by the spread of tomato powdery mildew (caused by *Oidium neolyccopersici*) since 1986 and the realization that all the cultivars being grown were susceptible. In contrast, many wild accessions in the genebanks in Davis, California and the Netherlands were resistant. Crosses were made with susceptible cultivar Moneymaker and F<sub>2</sub> populations produced for genetic studies on the inheritance of resistance and its association with molecular markers. Three monogenic *Ol*-genes were found (in *S. peruvianum* and *S. habrochaites*) as well as three QTLs for resistance (in *S. neorikii*). Introgression of the *Ol*-genes was done by backcrossing (at least three generations of up to 20 plants) with selection for resistance in a reliable seedling disease test, although molecular markers could have been used. Lines true breeding for resistance were obtained on selfing and those most similar to cultivar Moneymaker on the basis of AFLP markers were maintained as near-isogenic lines, and made available to private breeding companies. Bai (2012) concludes by saying that near-isogenic lines for the QTLs are being developed by marker assisted selection.

## New Rice for Africa and Linkage Drag in Rice

The genus *Oryza* has 22 species of which two are cultivated, *O. sativa* (Asian rice) and *O. glaberrima* (African rice). Success in securing a few fertile grains from 7 out of 48 crosses between these species (Jones et al. 1997), followed by two back-crosses to the *O. sativa* parents, resulted in the development of New Rice for Africa (NERICA). The products combined the yield potential and grain characteristics of *O. sativa* with the weed competitiveness and tolerance to drought and adverse soil conditions of *O. glaberrima* (Virmani and Ilyas-Ahmed 2007).

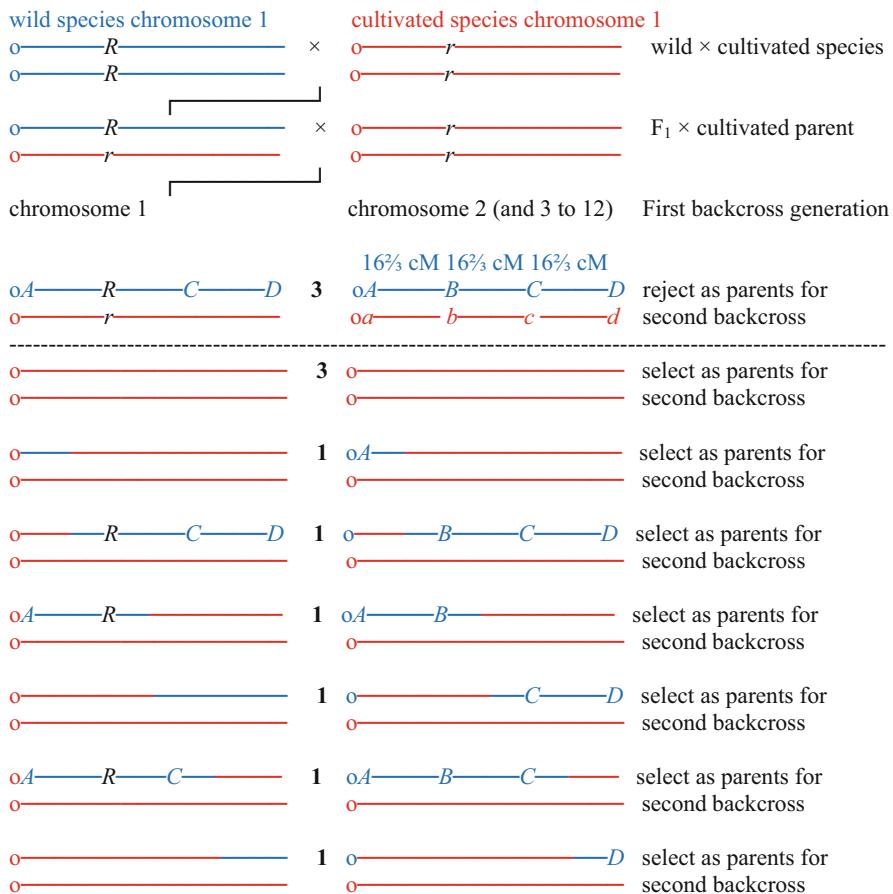
The wild relatives of *O. sativa* and *O. glaberrima* have been used in introgression breeding (Guimarães 2012), but here we are going to look at an example of the use of landraces and the extent of 'linkage drag'. Jia (2009) studied linkage drag in rice for the introgression of major dominant gene *Pi-ta* which confers resistance to AVR-Pita races of the filamentous ascomycete fungus *Magnaporthe oryzae*, the cause of rice blast disease. The resistant tropical japonica cultivar Katy, released in the 1990s, inherited *Pi-ta* from a Vietnamese indica landrace variety Tetep, and has since been used as donor of the gene in breeding the resistant US cultivars Kaybonnet, Drew, and Madison. In Japan, indica landrace Tadukan from the Philippines was used as the donor. Both Tetep and Tadukan were the progenitors of IR64, the world's most popular indica cultivar, but their genomic relationship is unknown. Jia (2009) backcrossed *Pi-ta* from cultivar Katy into susceptible temperate japonica cultivar M202 for five generations with selection each generation for resistance to AVR-Pita race IB49. Two progeny of each of the 22 BC<sub>5</sub>F<sub>1</sub> were genotyped using 12 simple sequence repeat markers around the *Pi-ta* genomic region on chromosome 12. A few recombinations had occurred at the distal ends of chromosome 12 but the majority of the central portion including the centromere

and *Pi-ta* had not recombined. Genomic fragments around *Pi-ta* ranged from half (14 Mbp) to the entire donor chromosome (27 Mbp), whereas SSR markers on chromosomes 1–11 in the 43 (one lost) BC<sub>5</sub>F<sub>2</sub> individuals were primarily from the recurrent parent M202. Similarly, large segments of comparable sizes of the *Pi-ta* genomic region originating from Tetep were also identified in *Pi-ta* containing cultivars, Katy, Kaybonnet, Drew and Madison. Furthermore, Tetep had an identical chromosome 12 to Tadukan, and IR64 was found to contain the same 6.4 Mbp around *Pi-ta*. In conclusion, this study revealed that a large segment of chromosome 12 was inherited over five backcross generations, and linked genes on chromosome 12 have not been recombined during the breeding of several elite rice cultivars from landraces. Further research is required to determine if selection of the linkage block was due to suppression of recombination or the presence of other desirable genes; but as the resulting cultivars were successful, the ‘linkage drag’ was not a problem, possibly because landraces were the source of resistance.

## Molecular Marker-Assisted Introgression

Molecular marker-assisted introgression offers the possibility of faster progress than can be achieved by traditional backcrossing. This is because one can select genotypically against the unadapted species genome as well as genotypically for the desired gene(s). With adequate molecular marker coverage of all chromosomes, it is possible to estimate the optimal combination of population sizes and number of backcross generations, and to select in a very precise way for the desired products of meiosis in each backcross generation (Hospital 2003).

It is now time to illustrate the concept of molecular marker-assisted introgression, but without going into the theory in detail. Let us consider the introgression (backcrossing) of a dominant gene (*R*) for disease resistance from a wild to a cultivated species where both are diploid species with 12 pairs of homologous chromosomes ( $2n = 2x = 24$ ), and there is always just one chiasma (crossover) per chromosome pair. Furthermore, this chiasma is equally likely to occur at any location along the chromosome pair. The resistance gene is on chromosome 1 at 16½ cM (Fig. 15.2). In the first backcross generation it is present on average in half of the offspring (*R:r* segregates 1:1). If phenotypic selection for resistance is 100 % effective (or a molecular marker occurs within the gene), one half of the selected offspring will have a whole chromosome 1 from the wild species. The other half of the progeny will have varying lengths of chromosome 1, depending on where the chiasma occurred. In subsequent backcrosses the length linked to the *R*-gene will reduce as a result of more crossovers, and some of these will be really close to the *R*-gene. Eventually the selected offspring will have very little of wild-species chromosome 1 attached to the resistance gene. Nevertheless, some undesirable wild-species alleles may be retained through their linkage to the selected gene, a phenomenon known as linkage drag. If crossovers do not occur in the region of the *R*-gene, then linkage drag can be more of a problem. Hospital (2003) looked at



**Fig. 15.2** Molecular marker assisted introgression (backcrossing) of resistance gene *R* (on chromosome 1) from wild to cultivated species: both diploid species with 12 pairs of homologous chromosomes ( $2n = 2x = 24$ ) and one chiasma per chromosome pair (in the first backcross generation there are eight types of offspring for four markers, depending on where single chiasma occurs)

linkage drag in detail and concluded that a typical marker-assisted backcross scheme should involve 3–4 backcross generations, rather than just one or two. This permits a more drastic reduction of linkage drag while reducing the genotyping effort, and increases the probability of success (obtaining a double recombinant) in advanced backcross generations.

Now let us consider a gene (*A/a* say) on chromosome 2. In the absence of selection, half the offspring in the first backcross generation will be *Aa* and half *aa* (*Aa*  $\times$  *aa* gives *Aa:aa* offspring in 1:1 ratio). Then in the absence of selection, one quarter of the offspring in the second backcross will be *Aa* and three quarters *aa*

( $Aa \times aa$  and  $aa \times aa$  combined give  $Aa:aa$  offspring in 1:3 ratio). Hence the frequency of the cultivated allele  $a$  increases with each backcross generation: 0.750, 0.875, 0.938, 0.969, 0.984, 0.992, etc. Generalizing we can say that on average, in the sixth backcross generation we will have  $Rr$  and  $rr$  genotypes in equal frequency (frequency of  $r = 0.75$ ), but otherwise the cultivated parent's alleles at a frequency greater than 99 %. If we are dealing with an inbreeding cultivated species, two generations of selfing will ensure that we have an inbred line that is true breeding for the resistance gene ( $Rr$  selfed gives  $RR$ ,  $Rr$  and  $rr$ ; and  $RR$  selfed will be true breeding), but otherwise resembles the cultivated parent apart from any effects of linkage drag.

In practice in the backcross generations some offspring will display more cultivated-species like phenotypes than others. Hence some phenotypic selection can be practised in favour of the cultivated-parent phenotype and against the wild-species phenotype. As a consequence, the cultivated allele frequencies will increase faster than by chance and fewer than six backcrosses may suffice. However, let us now consider the situation where we have whole genome coverage with equally spaced molecular markers; four (A, B, C and D) per chromosome in our simple example. Let us look at chromosome 2 in Fig. 15.2. The presence of the four alleles A to D in an individual tells us that it contains the whole chromosome 2 from the wild species (frequency  $\frac{1}{4}$ ). Hence  $\frac{3}{4}$ th of individuals in the first backcross do not have a whole chromosome 2 from the wild species, just a part of it or none of it. The same is true for chromosomes 3–12. Therefore  $(\frac{3}{4})^{11}$  ( $= 0.0422$ ) of the individuals lack whole chromosomes from the wild species for all 11 chromosomes. Furthermore, in selecting for the  $R$ -gene on chromosome 1, we can also use markers to eliminate individuals with a whole chromosome 1 from the wild species. In other words we select one quarter of individuals. The total proportion we select is therefore  $\frac{1}{4}(\frac{3}{4})^{11}$  ( $= 0.011$ ). This means that in a modest sized first backcross generation of 200 individuals, on average 2 can be found that harbour the  $R$ -gene but no whole chromosomes from the wild species. We have already fixed parts of all of the cultivated parent's chromosomes. The average frequency of the cultivated allele (Fig. 15.2) is 0.833 compared with 0.75 by chance. It should be pointed out that there is no chance of fixing all of the cultivated-parent apart from the  $R$ -gene in the first backcross generation because  $\frac{1}{4}(\frac{1}{4})^{11}$  is extremely small ( $5.96 \times 10^{-8}$ ).

The process of eliminating as many as possible of the wild-species alleles can be repeated in the second and third backcrosses. The results for the second backcross generation are shown in Table 15.1. One third of their parents had two copies of chromosome 2 and hence were true breeding (Fig. 15.2). The other two thirds produced one quarter of offspring with two copies. Hence half ( $1/3 + 2/3 \times \frac{1}{4} = \frac{1}{2}$ ) of the offspring have two copies of chromosome 2 from the cultivated parent as a result of no recombination. Furthermore, a total of 43 out of 54 offspring can be selected that have either no marker or just one marker from the wild species. The same is true for chromosomes 3–12. Therefore  $(43/54)^{11}$  ( $= 0.0816$ ) offspring can be selected that have either no marker or just one marker from the wild species for all 11 chromosomes. Furthermore, in selecting for the  $R$ -gene on chromosome 1,

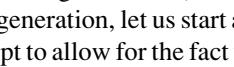
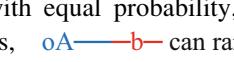
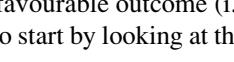
**Table 15.1** Second backcross generation: frequencies of marker genotypes on chromosomes without *R*-gene and chromosome with *R*-gene (*R*-gene replaces marker *B*), and genotypes selected as parents of third backcross (the *abcd* marker class can be divided into whole cultivated chromosomes that have not undergone recombination, whole ones that are the result of recombinations (estimated) and recombinants that may contain short lengths of wild species chromosomes)

Marker genotypes	Out of 54	Select (S)	<i>R</i> -gene	Out of 36	Select (S)
<i>abcd</i> (whole chromosome)	27	S			
<i>abcd</i> (whole recombinants)	3	S			
<i>abcd</i> (recombinants)	3	S			
<i>Abcd</i>	4	S			
<i>aBCD</i>	2		<i>aRCD</i>	4	
<i>ABcd</i>	3		<i>ARcd</i>	6	S
<i>abCD</i>	3				
<i>ABCd</i>	2		<i>ARCd</i>	4	
<i>abcD</i>	4	S			
<i>aBcd</i>	1	S	<i>aRcd</i>	2	S
<i>aBCd</i>	1		<i>ARCd</i>	2	S
<i>abCd</i>	1	S			

**Table 15.2** Frequencies of genotypes in parents and offspring of third backcross generation, for chromosomes without *R*-gene

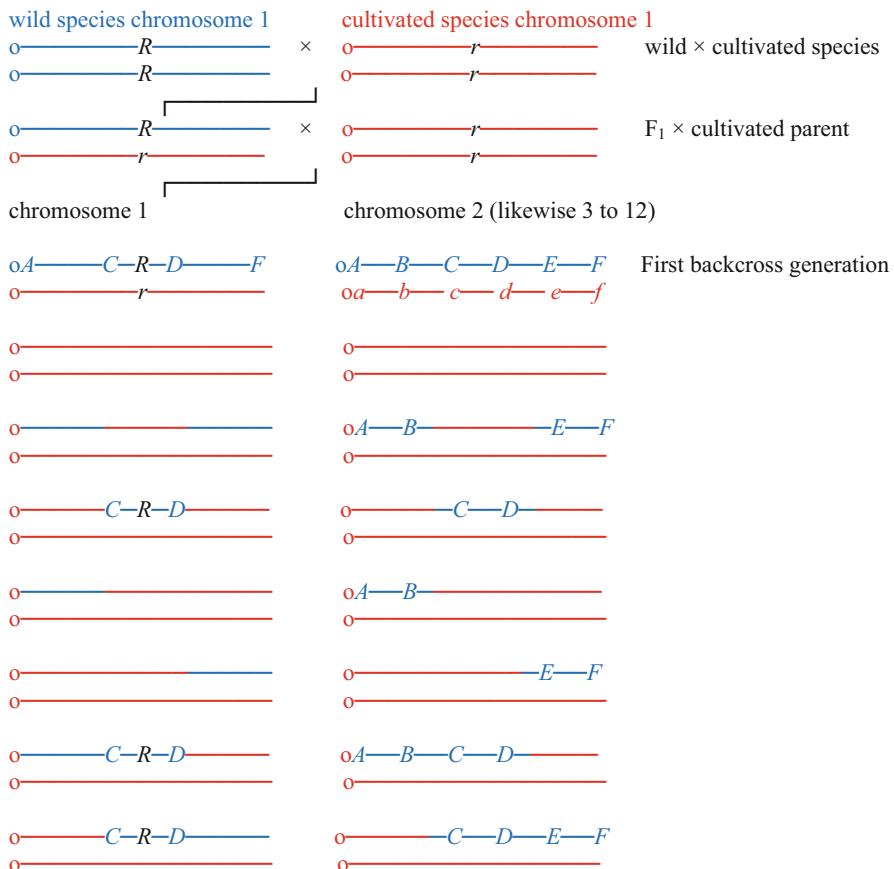
Parents	Frequency	Offspring
<i>abcd</i> (whole chromosome)	27/43	All <i>abcd</i> (whole chromosome)
<i>abcd</i> (whole recombinants)	3/43	All <i>abcd</i> (whole chromosome)
<i>abcd</i> (recombinants)	3/43	>1/4 <i>abcd</i> (whole chromosome), about 5/12
<i>Abcd</i>	4/43	>1/4 <i>abcd</i> (whole chromosome), about 5/12
<i>abcD</i>	4/43	>1/4 <i>abcd</i> (whole chromosome), about 5/12
<i>aBcd</i>	1/43	>1/4 <i>abcd</i> (whole chromosome), about 1/3
<i>abCd</i>	1/43	>1/4 <i>abcd</i> (whole chromosome), about 1/3

we can eliminate individuals with two markers from the wild species (*aRCD* and *ARCd*). In other words we select (5/18) of the individuals. The total proportion we select is therefore  $(5/18)(43/54)^{11}$  (= 0.0223). This means that in a modest sized second backcross generation of 200 individuals, on average 4–5 can be selected as parents of the third backcross generation (Table 15.2).

In the third backcross generation, let us start again by considering chromosome 2. Also let us make an attempt to allow for the fact that the chiasma can occur anywhere between two markers with equal probability, and not simply halfway between them. In other words,  can range from  to 

with marker genotype *abcd* (whole chromosome) will breed true and produce offspring with this genotype. It is also estimated that half of the remaining *abcd* parents will have whole chromosomes as a result of recombination, and hence breed true. Therefore 30/43 parents will produce offspring with two copies of chromosome 2 from the cultivated parent. Furthermore the remaining proportion of parents (13/43) will produce a quarter of such offspring because there is just one chiasma per chromosome pair (refer back to Fig. 15.2). Hence at least 133/172 (0.773) offspring will be fixed (two copies) for chromosome two from the cultivated parent. In addition, whole chromosomes from the cultivated parent will be reformed through recombination, and this contribution is estimated at 8/172, making a total of 141/172 (0.820). The same result is true for chromosomes 3–12. Therefore  $(0.820)^{11}$  ( $= 0.112$ ) offspring will be identical to the cultivated (recurrent) parent for chromosomes 2–12, and half of these individuals will carry the *R*-gene (0.0562), of which 4/9 (0.025) can be shown to have no markers from the wild species (i.e. lack A, C and D). Hence in a modest sized third backcross generation of 200 individuals, on average 5 will have secured our objective; namely a genotype identical to the cultivated parent, apart from an introgressed *R*-gene and some linked genes from the wild species on chromosome 1. In other words, in a modest sized (200 individuals per generation) backcrossing programme we have achieved our objective in just three generations; something that is virtually impossible in one generation.

We could go on to consider the situation where there are always two chiasmata (crossovers) per chromosome pair. The first backcross generation is shown in Fig. 15.3 for chromosomes 1 and 2, the latter having 6 markers. The presence of the six alleles *A* to *F* in an individual tells us that it contains the whole chromosome 2 from the wild species (frequency 1/8). Hence 7/8th of individuals in the first backcross do not have a whole chromosome 2 from the wild species, just a part of it or none of it. The same is true for chromosomes 3–12. Therefore  $(7/8)^{11}$  ( $= 0.2302$ ) of the individuals lack whole chromosomes from the wild species for all 11 chromosomes. Furthermore, in selecting for the *R*-gene on chromosome 1 (in centre of chromosome), we can eliminate individuals with a whole chromosome 1 from the wild species. In other words we select 3/8th of individuals. The total proportion we select is therefore  $3/8(7/8)^{11}$  ( $= 0.0863$ ). This means that in a modest sized first backcross generation of 200 individuals, on average 17 can be found that harbour the *R*-gene but no whole chromosomes from the wild species. In other words, we have already fixed parts of all of the cultivated species chromosomes. The average frequency of the cultivated allele (Fig. 15.3) is 0.7857 compared with 0.75 by chance. Again there is no chance of fixing all of the cultivated-parent apart from the *R*-gene;  $3/8(1/8)^{11}$  is extremely small ( $4.366 \times 10^{-11}$ ). The process of eliminating as many as possible of the wild-species alleles can be repeated in the second and third backcrosses, by which time most of them will have disappeared. The details are more complicated now that there are two chiasmata per chromosome and are not shown. The general theory can be found in the account by Hospital (2003) and the process explored by computer simulations with realistic chiasmata frequencies and distributions for any given crop species.



**Fig. 15.3** Molecular marker assisted introgression (backcrossing) of resistance gene *R* (on chromosome 1) from wild to cultivated species: both diploid species with 12 pairs of homologous chromosomes ( $2n=2x=24$ ) and two chiasmata per chromosome pair (there are eight equally frequent types of offspring in first backcross generation)

Hospital (2003) considered both reduction of linkage drag on the carrier (*R*-gene) chromosome and selection on non-carrier chromosomes. He concluded that dense coverage of the non-carrier chromosomes by molecular markers is not necessary to increase the overall recipient genome content; in fact four markers are sufficient for a chromosome of 100 cM (b-e in Fig. 15.3). Furthermore, three or four backcross generations are sufficient to increase the recipient genome content on non-carrier chromosomes above 99 %, compared with six backcross generations without selection. Hospital (2003) considered the combined selection in three steps: foreground selection of all individuals heterozygous at the target locus (*R/r*), reduction of linkage drag by selection of all individuals that are homozygous for the recipient

allele at one or preferably both markers flanking the target locus ( $ccRrDD$ ,  $CCRrdd$  or preferably  $ccRrdd$ ), and background selection on non-carrier chromosomes for the individual that is most homozygous for the most markers on non-carrier chromosomes ( $aabbccdd$ ). Target-marker distance on the carrier chromosome was 2 cM and each non-carrier chromosome (100 cM) was controlled by three markers located at optimally determined positions. The results of simulations, averaged over 1000 replicates, were 99 % recipient genome content achieved in four backcross generations, compared with six backcross generations without selection. All selected markers had returned to homozygosity of recipient type, both for the carrier chromosome and the non-carrier ones. The population sizes in each generation were optimized over the four backcross generations and the results were 70, 100, 150 and 300. If maximal efficiency was sought in only three generations, then larger population sizes (117, 171 and 370) would be required to provide 98.5 % recipient genome content in backcross three. The total of 658 individuals is not that different from the three backcrosses of 200 individuals (total 600) in my simple example. One final point from Hospital (2003) is worth making here. It is not possible to simultaneously manipulate more than five or six known target genes in a marker-assisted backcross programme because of the population sizes required.

Lecomte et al. (2004) demonstrated that the kind of backcrossing programme just considered can be used in tomato to transfer into elite lines, five chromosomal regions carrying important QTLs involved in fruit quality. The regions were located on chromosomes 1, 2, 4 and two regions of 9. The researchers produced 144 recombinant inbred lines (RILs) from a cross between a cherry tomato (Cervil) chosen for its good flavour and a line (Leovil) with bigger but less tasty fruits. QTL analysis revealed that one RIL had favourable Cervil alleles in all five regions. It was used as the donor parent for three marker-assisted backcrossing programmes with recipient lines Leovil, Vil B and Vil D provided by the Vilmorin seed company. The first crosses between recipients and donor were genetically homogeneous but called the BC<sub>1</sub> (rather than F<sub>1</sub>) generations. Almost 300 plants were grown per recipient line in BC<sub>2</sub> and marker-assisted selection used to select a single plant as parent of the BC<sub>3</sub> generation. The process was repeated and a single plant out of 300 selected from BC<sub>3</sub>, but this time two selfing generations (300 and 40 individuals, respectively) were produced to fix the Cervil alleles at all five QTL regions. Thus three improved BC<sub>3</sub>S<sub>2</sub> lines in total were obtained with homozygous alleles at the five regions. The researchers used these lines for further genetic analysis, the results of which can be found in their paper.

## Chromosome Manipulation (Engineering) in Wheat

Genera related to wheat (*Triticum*) such as *Secale*, *Agropyron*, *Aegilops*, *Haynaldia*, *Elymus*, *Thinopyrum* and *Hordeum* represent an enormous pool of genetic variation that is potentially available for wheat improvement; but only if barriers to crossing, chromosome pairing and recombination between

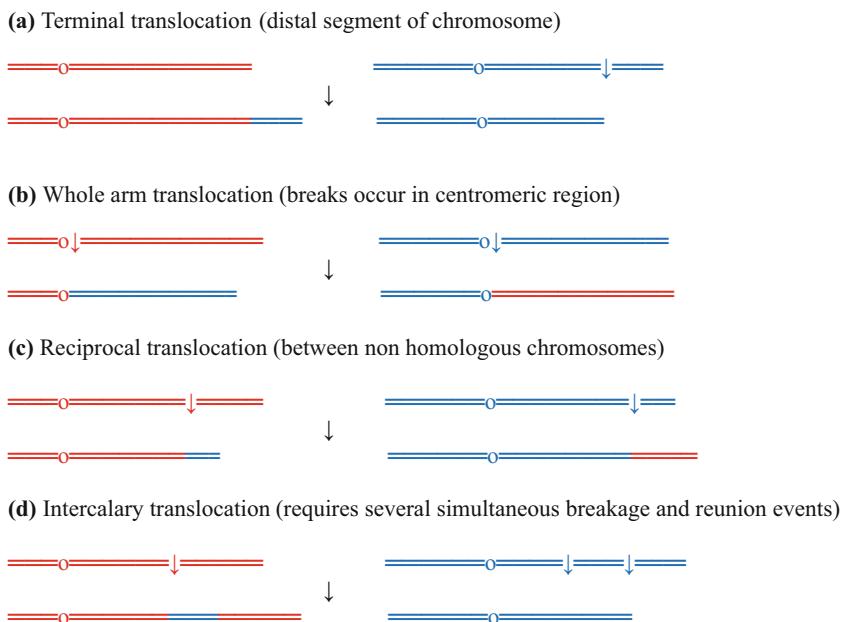
homoeologous (distantly related) chromosomes from different genomes can be overcome (Wang et al. 2012). The gene transfer to cultivated wheat can be achieved by chromosome manipulation involving the development of amphiploids with a complete set of the alien chromosomes and then alien chromosome addition or substitution lines by backcrossing (AABBDD + xx or AABBDD – aa + xx, AABBDD – bb + xx, AABBDD – dd + xx, where x is alien chromosome). The final incorporation of genetic material from addition or substitution lines depends either upon the management of homoeologous pairing, which is under genetic control, or upon the restructuring of chromosomes by translocation. The result will be a translocation line with a small alien chromosome segment in an otherwise normal wheat chromosome. Translocation lines are genetically fixed and stable and can be used in traditional breeding.

### ***Management of Homoeologous Pairing***

Chromosome pairing in regular allopolyploids is only between homologous chromosomes within genomes (e.g., in wheat, A with A, B with B, but not A chromosomes with B chromosomes). However, the constituent genomes of allopolyploids have extensive homology from earlier evolutionary identity (i.e. identity by descent). Genomes and individual chromosomes that are so related are said to be homoeologous (for wheat, for example, the A and B chromosomes from the A and B genomes). The functional diploidy and success of all well established allopolyploids, shows that some sort of homoeologous pairing suppression usually evolves. In wheat, for example, a dominant gene (*Ph1*) on the long arm of chromosome 5B stabilises polyploidy by both promoting homologue pairing and preventing MLH1 (a DNA mismatch repair protein) sites from becoming crossovers on paired homoeologues during meiosis (Martín et al. 2014). Since the absence or mutation of *Ph1* allows homoeologous pairing, manipulation of this gene has been a major approach for transferring desirable genes from related species to common wheat (*T. aestivum*, AABBDD). The classic example is the 5B-suppressing effect of *Ae. speltoides* (SS, similar to BB) that was used to effect the transfer of yellow rust resistance from *Aegilops comosa* (MM) into wheat (Riley et al. 1968). Two sets of backcrosses were involved. The first added a single chromosome (2M) carrying resistance gene *Yr<sub>8</sub>* from *Aegilops comosa* to *T. aestivum* cultivar Chinese Spring, thus creating an addition line (AABBDD + 2M). Then a hybrid (ABDS + 2M) was made with *Ae. speltoides* to encourage homoeologous pairing during backcrossing to cultivar Chinese Spring, with selection for resistance. The result was insertion of a segment of 2M carrying *Yr<sub>8</sub>* into a 2D chromosome of Chinese Spring. After two generations of selfing, cultivar Compair was produced with a full complement of 42 chromosomes (AABBD'D'), including two '2D/M' chromosomes homozygous for *Yr<sub>8</sub>*. Since then a large number of lines containing alien genes for resistance to abiotic and biotic stresses have been developed using allele *ph1b* (Mergoum et al. 2009a).

## Translocation Breeding

Translocations were briefly introduced in Chap. 3, with a wheat example, in the context of naturally occurring mutations. In the next chapter on mutation breeding we shall see that radiation, such as X-rays and gamma rays, breaks chromosomes more or less randomly and that the broken ends heal, rejoin or fuse in new combinations. In the current context of translocation breeding, the last possibility allows the formation of new chromosomes carrying translocated fragments of the foreign chromosome. If a translocated fragment contains the desired gene but no other genes with undesirable effects, and there are no undesirable position effects resulting from genes in new chromosome locations, the new recombinant chromosome can be incorporated in the genome of the recurrent parent. This usually involves a lot of screening to select such a recombinant chromosome. Four basic types of translocation can occur, as shown in Fig. 15.4. Friebe et al. (1996) used C-banding and *in situ* hybridization analyses to identify wheat and alien chromosomes involved in wheat-alien translocations and to determine their size and breakpoint positions. Out of the 57 spontaneous and induced wheat-alien translocations, 45 were terminal translocations with distal alien segments translocated to



**Fig. 15.4** Types of translocation: cultivated chromosome in *red* and wild (alien) one in *blue*, with break points shown by *red* and *blue* arrows and centromeres by *circles*. (a) Terminal translocation (distal segment of chromosome). (b) Whole arm translocation (breaks occur in centromeric region). (c) Reciprocal translocation (between non homologous chromosomes). (d) Intercalary translocation (requires several simultaneous breakage and reunion events)

wheat chromosome arms, ten were whole arm translocations with the breakpoints within the centromeric regions and just two were intercalary wheat-alien translocations. Wang et al. (2012) explain in detail the steps involved in translocation breeding: identification of the wild species with the desired trait and hence gene(s); production of the wheat-wild species hybrid (chromosomally doubled amphiploid) and alien addition or substitution lines; irradiation of these lines (or hybrid itself if mass of translocation lines involving different chromosomes required) to induce translocations; identification and selection of desired translocation events (phenotypic, cytological and molecular); and finally fixation of the translocated chromosome by selfing to produce homozygous lines.

The classic example is the transfer by Sears (1956) of resistance (gene *Lr9*) to leaf rust (*Puccinia triticina*) from *Aegilops umbellulata* (now classified as *Triticum umbellulata*) to *Triticum aestivum* wheat. The key steps were (a) the bridging cross between *Triticum turgidum* ssp. *dicoccoides* (genomes A and B) and *Ae. umbellulata* (U genome) which on chromosome doubling gave the allohexaploid AABBUU, (b) the cross between this allohexaploid and *T. aestivum* (A, B and D genomes) which gave the hybrid AABBUD in which 14 bivalents and 14 univalents formed during meiosis, (c) two backcrosses to *T. aestivum* using the latter as female parent and selection of a resistant plant that was AABBDD plus one *Ae. umbellulata* chromosome, (d) selection of a resistant plant in the progeny with an isochromosome that was duplicated for the long arm of the *Ae. umbellulata* chromosome, (e) treatment of plants carrying the isochromosome with X-rays and use of their pollen on *T. aestivum* and (f) identification of 40 translocations in 132 resistant plants out of 6091, of which one was assumed to be an intercalary translocation that was cytologically undetectable because it formed 21 normal bivalents in meiosis. In the intercalary translocation a small portion of the *Ae. umbellulata* chromosome had been removed as a result of two double strand breaks and inserted in a wheat chromosome as a result of one or two double strand breaks. Further study showed that a reciprocal exchange had in fact occurred through a terminal transfer of the long arm of the 6U segment to the distal region of the long arm of 6B (Friebe et al. 1996). The translocation line was used as a source of resistance to leaf rust in breeding several commercially successful cultivars in the USA (Wang et al. 2012).

Wang et al. (2012) cite two subsequent commercial successes out of the many translocation lines that have been produced for wheat improvement. Firstly, the wheat-rye 1RS/1BL (homoeologous chromosome 1, short-arm of R and long-arm of B genome) translocation has been used successfully in breeding programmes worldwide because rye 1RS contains several useful disease resistance genes. Secondly, since 1971 a total of 12 Australian wheat cultivars have stem rust resistant gene *Sr26* as a result of the transfer of a segment of *Agropyron elongatum* chromosome 6E to chromosome arm 6AL of wheat.

Both Friebe et al. (1996) and Wang et al. (2012) reached the same conclusions about the relative merits of management of homoeologous pairing and restructuring of chromosomes by translocation. The majority of translocations resulting from radiation treatments were formed between nonhomoeologous chromosome arms

and were therefore genetically unbalanced, leading to reduced agronomic performance. Hence strong selection is required for the recovery of compensating translocations. In contrast, all wheat-alien translocations produced by induced homoeologous recombination are of a compensating type and therefore have greater agronomic potential. However, recombination between homoeologous chromosomes of wheat and related species is drastically reduced in the proximal regions of chromosome arms, and this is a limiting factor.

## Base Broadening by Incorporation

As mentioned in the introduction to this chapter, since the 1960s there has been interest in a more extensive broadening of the genetic base of crop production, described by some as incorporation to distinguish it from introgression of specific genes for specific traits. Many examples can be found in the book *Broadening the Genetic Base of Crop Production* edited by Cooper et al. (2001), and further examples in the *Handbook of Plant Breeding* (Prohens et al. 2008). The general genetic features of incorporation are use of the broadest possible starting material, extensive recombination from natural or controlled hybridization, weak selection progressively decentralized to target environments, and sufficient time to produce improved material that can be crossed with elite germplasm without adverse effects on yield and agronomic performance because it is adapted to those environments. We are looking in a different way at the long-term response to population improvement from a broad-based foundation population dealt with in Chap. 10, and the composite crosses dealt with in Chap. 13. The outcome of an effective base-broadening programme is enhanced genetic variance in economic characters and either good material *per se* or good parents for crossing into established programmes. The 15 International (CGIAR) Research Centres (Chap. 2) play their role by making improved germplasm available to individual developing countries for further breeding through National Agricultural Research Systems (NARS). The interested reader is referred to their websites. Here we are going to look at base broadening in two crops, potato and maize, that included changes in adaptation from short to long days.

## Base Broadening of Potatoes

### *Base Broadening of Potatoes in Europe and North America*

During the second half of the twentieth century recognition was given to the value of the cultivated species of South America for broadening the genetic base of European and North American breeding programmes. One of the earliest examples

of base-broadening was the production of Neotuberosum (or long-day Andigena) potatoes, but there are different views on its success. As a potato breeder and geneticist I have tried to give fair assessments of the programmes in my reviews of potato breeding, the most recent of which was published in 2010 (Bradshaw and Bonierbale 2010). The first Neotuberosum programme was started by Simmonds from the Commonwealth Potato Collection in 1959 (Simmonds 1969). A gene pool of (short-day) Andigena potatoes with origins approximately 45 % Bolivian accessions from the CPC, 35 % south Peruvian, 10 % north Peruvian and 10 % Colombian was subjected to recurrent mass selection on a 2-year cycle in outdoor plots in Great Britain (northern latitude, long-day summer conditions). In the first year large populations of seedlings were grown in the field and selected for high yields of tubers of acceptable sizes, shapes and colours. In the second year the selected tubers were planted in an isolation site and open-pollinated berries harvested to provide the next generation of seedlings. Within four generations, the better Andigena clones were comparable in yield and maturity to (long-day) Tuberosum cultivars and better on average in terms of late blight resistance. These clones were more variable in tuber shape than modern cultivars and inferior in regularity of tuber shape, but of similar cooking quality. As a consequence of their rather 'rough' appearance, it subsequently proved difficult to breed successful cultivars from crosses of this Neotuberosum material with intensively selected Tuberosum clones, despite yield heterosis in crosses to modern cultivars. Just one cultivar *Shelagh* (released 1986), out of 50 released from 1968 to 2008 by the Scottish Plant Breeding Station and Scottish Crop Research Institute (Bradshaw 2009a), had a Neotuberosum parent despite such parents being used over the period from 1969 to 1998, with intensive evaluation of hybrids from 1979 to 1987. Neotuberosum needed improving for traits in addition to tubering in long days to have real impact because it was not as good as intensively selected Tuberosum. However, recurrent mass selection was discontinued in 1978, when the emphasis changed to evaluating the potential of this improved material in breeding cultivars, a task that was essentially completed by 1987. In contrast, 15 out of the 50 cultivars starting with *Pentland Javelin* in 1968 have the *H1* gene for resistance to *Globodera rostochiensis* introgressed from a short-day Andigena accession, as does the most widely grown cultivar in Great Britain, *Maris Piper* released in 1967. The original introgressions required just three backcross generations.

Experiences were similar elsewhere in Europe and in North America, as reviewed by Bradshaw and Bonierbale (2010). In the USA Plaisted (1987) introduced deliberate pollination of selected clones with bulk pollen to avoid inbreeding depression from selfing in his Neotuberosum programme. He also achieved more vigorous plants for selection by raising his seedlings in pots to produce seed tubers for planting field trials. This enabled him to screen the seedlings in the glasshouse for resistance to late blight and viruses, but it also increased the length of each cycle to 3 years. By 1987 his programme had released just one cultivar, *Rosa*, which was a *Tuberosum* × *Neotuberosum* hybrid. In Canada, Tarn and Tai (1983) concluded that first-generation hybrids between *Tuberosum* and *Neotuberosum* were not the best use of their *Neotuberosum*. They also advocated further improvement of their

Neotuberosum and/or backcrosses of the hybrids to Tuberosum. No doubt the debates will continue about the relative merits of introgression and incorporation in potato.

### **CIP Populations B1 to B3**

On its establishment, CIP recognized the need to make broad-based germplasm and candidate varieties from the world collection available to National Programmes in developing countries, particularly germplasm with durable resistance to late blight. Probably the best known broad-based germplasm from CIP's breeders are the *R*-gene (race-specific) free B populations with quantitative resistance to late blight. B1 was a population of short-day-adapted Andigena potatoes, B2 was derived from crosses of the Andigena clones with long-day-adapted Tuberosum clones lacking *R*-genes, and B3 was a population which in addition contained resistances from wild species but from which *R*-genes had been removed by progeny testing. The aim was to help with the development of improved cultivars, for a wide range of environments, which possessed high and stable levels of resistance to late blight in combination with resistances to viruses and suitable tuber type, and culinary and processing quality (Trognitz et al. 2001; Landeo 2002). The population improvement involved recurrent selection for quantitative resistance to late blight and other desirable traits under high endemic disease pressure in the Andean highlands together with selection in those geographical areas where the new cultivars were to be grown, for example in short days in East Africa (Mulema et al. 2004). Evaluation under long days in Argentina (Trognitz et al. 2001) and in South Korea was included to broaden the range of adaptation of the largely short-day adapted population. Clones with good general combining ability have been identified for use as parents in the local breeding programmes of the National Agricultural Research Systems in developing countries. Landeo (2002) reported good progress in selecting for resistance and other traits in the B populations, large additive genetic variances and high heritability estimates for resistance, and stability of resistance across diverse environments and pathogen populations in tropical environments. Bradshaw and Bonierbale (2010) also cover other base-broadening work at CIP and the Central Potato Research Institute of India.

### **Germplasm Enhancement of Maize**

Maize (*Zea mays*,  $2n = 2x = 20$ ) was domesticated from annual teosinte (*Zea mays* ssp. *parviglumis*) in the mid- to lowland regions of southwest Mexico around 9000 years ago (van Heerwaarden et al. 2011). The crop spread widely throughout the Americas both before and after Columbus discovered the New World in 1492. After 1492 maize was introduced into Europe, Africa and Asia and became a truly

global crop as we have seen in earlier chapters. Selection by indigenous farmers in all of these continents resulted in numerous locally adapted landraces, and by 1900, distinctive open-pollinated cultivars were also being maintained. As explained by Tallury and Goodman (2001), the tremendous amount of morphological variability which could be seen allowed cultivated maize to be classified and described in terms of about 250 races. These races, together with annual and perennial teosinte, make up the primary gene pool of maize. The secondary gene pool comprises perennial *Tripsacum* species ( $2n = 2x = 18$ , and polyploids), including obligate apomicts, and the tertiary gene pool related oriental or Asiatic genera. However, Tallury and Goodman (2001) thought that the 250 races would provide adequate diversity for future commercial exploitation, particularly as just one race, Corn Belt Dent, represented the commercial germplasm grown in the entire USA, and almost all of the temperate regions of the world. They also pointed out that since the advent of hybrid maize in the USA in the 1930s, most of the older open-pollinated cultivars have been replaced by hybrid ones in developed countries, whereas landraces and open-pollinated cultivars are still being grown by a substantial number of farmers in developing countries. The elite inbred lines that came to dominate hybrid production trace back to relatively few open-pollinated cultivars.

Hallauer and Carena (2009) explain that a consequence of the success of hybrid breeding in the USA was that the period 1920–1950 saw little general development of germplasm resources. Although some breeders and researchers argued that about 98 % of the world's maize germplasm was being ignored, prebreeding research programmes from 1950 to 2000 depended on those individuals who appreciated the possible contribution of exotic (unadapted) germplasm to modern-day maize breeding programmes. The results of specific studies can be found in the review by Hallauer and Carena (2009). The use of exotic tropical and subtropical germplasm in breeding programmes for temperate regions requires selection to reduce the effects of photoperiodism. Although this can be achieved relatively quickly, as we saw in Chap. 7, further selection of adapted tropical materials is required to meet current standards in hybrid cultivars for grain yield, root and stalk strength, pest tolerance and maturity. Furthermore, there is the question of the most appropriate proportion of tropical germplasm in the material being evaluated: the 50 % proportion in crosses with temperate germplasm can be reduced before selection to 25 or 12.5 % by one or two generations of backcrossing to elite temperate germplasm. Hallauer and Carena (2009) concluded that despite these efforts, sustained long-term programmes involving exotic germplasm were not continued because of inadequate financial support. They hoped that this may be rectified in the USA by the Germplasm Enhancement of Maize (GEM) project started in 1993.

## ***Germplasm Enhancement of Maize Project***

The Germplasm Enhancement of Maize (GEM) Project (<http://www.public.iastate.edu/~usda-gem/>) is a consortium of public and private organizations formed in 1993 to identify, improve, and develop a broad array of germplasm for use in present-day maize breeding programmes. Its ultimate aim is improving and broadening the genetic base of maize hybrids grown by American farmers. It has received substantial funding from the federal government since 1995. An account of how the project was set up and how it is organized has been given by Pollak and Salhuana (2001), and provides insight into public and private sector collaboration. Here we will concentrate on the scientific and technical aspects. The project was based on information and germplasm derived from the Latin American Maize Project (LAMP) which characterized and regenerated the maize accessions held in Latin American and US genebanks. In 1991 a catalogue and CD-ROM were published of data on 12,113 accessions evaluated in LAMP's first stage, and 2794 selected (primarily on yield) accessions evaluated in the second stage in 59 locations in Latin America and the USA. A total of 268 elite accessions were chosen and crossed with 31 testers and the data published in 1995 followed by a final report in 1997. However, it was already clear that further enhancement of this germplasm was required for it to enter commercial maize breeding programmes in the USA. The GEM project was initiated in 1994 with 51 elite tropical and temperate LAMP accessions chosen on the basis of yield performance in LAMP, together with seven commercial tropical hybrids. The hybridization protocol was for one of the private co-operating companies to cross an exotic material by a propriety inbred line to make a 50 % exotic breeding cross, then for another private co-operator to cross this with their proprietary line of the same heterotic group to create a 25 % exotic breeding cross. In 1996, 564 crosses (both 50 and 25 %) were evaluated for yield as testcrosses (tester from different heterotic group) at 4–9 locations, and the best selected for inbred line development by co-operators. A modified pedigree method was used in which early generation testing started at the S<sub>2</sub> generation (first-year trials) and was followed by further testing at the S<sub>3</sub> generation (second-year trials). The testers were elite proprietary inbred lines. The GEM project developed into a balance of developing and evaluating new breeding crosses plus ongoing line development at various stages. Pollak and Salhuana (2001) reported that over 350 breeding crosses had been developed and pedigree breeding for line development had been started in over 46 of these crosses. Progress can be found in the GEM Annual Reports which are posted on the project's website. The results demonstrate that exotic germplasm contains useful genes for yield, disease and pest resistance, and value added traits. GEM inbred lines considered of value for maize breeding are made available to US public sector programmes as well as to participating private companies. The new sources of germplasm are also available to all researchers free of charge through the North Central Regional Plant Introduction Station (NCRPIS). An example of research into breeding methods for GEM germplasm follows.

Jumbo et al. (2011) compared four breeding methods for each of three GEM breeding crosses. The crosses were (ANTIG01 × N16) × DE4, (AR16035 × S02) × S09 and (DKXL212 × S09) × S43b. ANTIG01 is a yellow semi-dent tropical Criollo race from Antigua, AR16035 is an orange flint temperate Cristalino Colorado race from Argentina, and DKXL212 is a Dekalb (Monsanto Company, St. Louis, MO) tropical single cross hybrid from Brazil. N16 is an elite non-Stiff Stalk line from a private co-operator, DE4 is a University of Delaware (Newark, DE) line derived from GEM breeding cross DKXL212 × N11a, S02, S09, and S43b are elite Stiff Stalk lines provided to GEM by unnamed private co-operators. The four breeding methods were the standard GEM modified pedigree method (CG), conventional mass selection (CM), modified single seed descent (MSSD) and doubled haploids (DH). The haploids for chemical doubling were produced by fertilization of F<sub>1</sub> plants with pollen from special haploid inducer stocks that results in subsequent paternal chromosome elimination and haploid seed production. The results revealed the theoretical and practical advantages and disadvantages of the four breeding methods, and hence will help inform future choices, but did not identify a best method for producing inbred lines. Perhaps this was because the differences between the methods were not really that large. The first segregating generation of each cross was called S<sub>0</sub>. All breeding methods had approximately 250 S<sub>1</sub> (selfed) families or (DH) lines per cross, those in the CM and MSSD schemes having been produced from individual plants selected from 1500. Fifty S<sub>2</sub> testcrosses (S<sub>2</sub> line × one tester) were evaluated for yield at 4–6 locations and 14 S<sub>3</sub> testcrosses (S<sub>3</sub> line × two testers) at 12 locations. Winter isolation blocks for crossing and summer yield trials for evaluation allowed completion in 3 years, with 14 lines compared by testcrossing from each breeding method and cross combination.

### ***Historical Genomics of North American Maize***

Genomic analysis of historical and contemporary germplasm is giving new insights into the genetic changes brought about by modern plant breeding. It therefore seems appropriate to end this section on maize with the results obtained by van Heerwaarden et al. (2012). They used high-density SNP genotyping (46,000 SNPs) to analyze 400 accessions of maize from the US Department of Agriculture (USDA)'s National Plant Germplasm System and other collaborators: 99 classic North America landraces (era 0), 94 early inbreds from before the 1950s (era 1), 70 advanced public lines from the 1960s and 70s (era 2) and 137 elite commercial lines from the 1980s and 90s (era 3) that were no longer under plant variety protection (ex-PVP). They found that the genomic history of maize was marked by a steady increase in genetic differentiation and linkage disequilibrium, whereas allele frequencies in the total population remained relatively constant. These changes were associated with increasing genetic separation of breeding pools (divergence of allele frequencies) and decreased diversity in the ancestry of individual lines. The three modern heterotic groups (Iowa Stiff Stalk Synthetic (SS),

Non-Stiff Stalk (NSS) and Iodent (IDT) were the result of ongoing divergence from a relatively homogeneous landrace population, with Yellow Dent as the main ancestral contributor to all three heterotic groups in the elite commercial lines. Nevertheless, small differences in contribution of Yellow Dent and Lancaster to SS and NSS lent some justification to the traditional distinction between these two heterotic groups on the basis of their landrace ancestry. Interestingly in era 2, three historically important breeding lines for elite germplasm were placed at the apices of the heterotic clusters: B73 in SS, Mo17 in NSS and 207 in IDT. The authors used linkage disequilibrium to characterize signals of directional selection at SNPs throughout the genome and identified a number of candidate genes of potential agronomic relevance. However, overall they found that only 5 % of SNPs showed some evidence of consistent selection with little evidence for individual lines contributing disproportionately to the accumulation of favourable alleles in today's elite germplasm. They concluded that maize breeding has involved a steady accumulation of changes at multiple loci through the selection and recombination of relatively common alleles, combined with heterosis due to differentiation of breeding pools. Furthermore, that selected traits of agronomic importance are predominantly quantitative in nature, with relatively few dominant contributions from individual alleles or lines.

## Participatory Plant Breeding and Diversity

One suggested feature of base broadening by incorporation is weak selection progressively decentralized to target environments. The latter is a feature of participatory plant breeding (PPB) as we saw in the rice example in Chap. 13. PPB is seen as a way of offering poor farmers in marginal areas in developing countries the possibility to decide which cultivars better suit their needs and conditions, without exposing them to any risk during the generations of selection. Ceccarelli and Grando (2009) have outlined a flexible model of PPB for inbreeding cereals in dry areas served by the International Centre for Agricultural Research in the Dry Areas (ICARDA), such as Algeria, Egypt, Eritrea, Iran, Jordan and Syria. Genetic variability is generated by the breeders, selection is conducted jointly by breeders, farmers and extension specialists in a number of target environments on-farm, and the best selections are used in further cycles of recombination and selection. The breeding method is a bulk-pedigree method. Crosses are made and the F<sub>1</sub> and F<sub>2</sub> generations grown in a field station before the F<sub>3</sub> to F<sub>6</sub> bulks of the crosses are yield tested over a period of 4 years in farmers' fields in villages in a number of countries, in initial (FIT), advanced (FAT), elite (FET) and large-scale (LS) trials. The decisions, on which bulks to retain for the following season, are made jointly by breeders and farmers based on both quantitative data and visual scores. There is the option, in countries with strict uniformity requirements for new cultivars, to extract pure lines from the best bulks of the previous cycle at the field station. The process can be independently implemented in a large number of

locations. Farmers handle the first phases of seed multiplication of promising breeding material in village-based seed production systems. As a result, cultivars reach the release phase earlier than in conventional breeding and are known to be acceptable to farmers. Another consequence is that biodiversity increases because different cultivars are selected in different locations, and hence the genetic base of future breeding is not narrowed as a consequence of just a few cultivars being grown and maintained.

# **Chapter 16**

## **Mutation Breeding**

### **Introduction**

W.K. von Röntgen discovered X-rays in 1895. Thirty two years later, in 1927, Hermann Joseph Muller (1927) presented proof that they could induce mutations within genes in the fruit fly *Drosophila melanogaster*. The heavy treatment caused a rise of about 15,000 % over that in the untreated germ cells. The same order of magnitude of dominant and recessive lethals occurred and these greatly outnumbered the non-lethals producing a visible morphological abnormality. Among the visible mutations found, the vast majority were recessive, although dominant ones did occur. In addition to the gene mutations, the X-ray treatment caused a high proportion of chromosomal rearrangements. Muller concluded his paper with the hope that the method presented would ultimately prove useful to the practical breeder, who had been compelled to remain content with the mere making of recombinations of the material already at hand, providentially supplemented, on rare and isolated occasions, by an unexpected mutational windfall! Spontaneous (natural) mutations arise at the very low rate of about one in one hundred thousand to one in a million per generation for most loci. Lewis John Stadler, who published his results on maize and barley a year after Muller in 1928, was the first to prove the feasibility of inducing mutations in higher plants; but he was a plant geneticist who was sceptical about their benefit to practical plant breeding (Stadler 1928a, b). This was because he thought that recombination would continue to provide plenty of useful variation whereas most newly induced mutations would be deleterious or strongly linked with other deleterious mutations. Nevertheless, his research laid the foundations for successful mutation breeding in plants. He showed that dry seed can withstand 15–20 times higher doses of radiation than germinating seed but that irradiation of the latter gave eight times higher mutation frequencies. Furthermore, he found that the mutation rate was proportional to the treatment doses (van Harten 1998).

Three firsts discussed by van Harten (1998) are worth a mention. The first mutant cultivar 'Chlorina' was obtained in tobacco after X-irradiation of inflorescences by Tollenaar and commercialized during the period 1934–1938. It had an extra chromosome and a gene mutation for a yellow-green leaf colour. The first example of the induction of mutations for disease resistance came in 1942 from the large scale mutation breeding programme of Freisleben and Lein in barley. They used X-rays to produce a recessive mutant (*ml-o 1*) for resistance to powdery mildew which proved highly effective and durable. Finally, the first example of the use of radiation-induced translocations in plant breeding was the use of X-rays by Sears in 1956 to transfer resistance to leaf rust from *Aegilops umbellulata* (now *Triticum umbellulata*) to wheat (*Triticum aestivum*). Since the 1960s, the most favoured physical mutagen has been gamma-rays, discovered by P. Villard in 1900.

Attempts to obtain artificially induced mutations by treating plant seeds with different chemicals started early in the twentieth century, but the first reports of real success came in the 1940s (van Harten 1998). For example, Gustafsson and MacKey (1948) proved that mustard gas was mutagenic in barley. However, being a bifunctional alkylating agent (has two reactive groups and hence can link bases) it was particularly disruptive in its effects. Hundreds of chemicals were subsequently tested, but mutation rates remained low because the toxicity of the mutagens prevented treatment with high concentrations. The most significant development was around 1960 when a number of researchers tested the mutagenic effect of ethyl methane sulphonate (EMS) on crop plants. This monofunctional alkylating agent (has a single ethyl group to donate) went on to become the most frequently and universally used chemical mutagen in plant breeding (van Harten 1998).

Since 1960 assessments of the possibilities and limitations of mutation breeding have taken place in many crop species. These have been helped by the establishment in 1964 of internationally co-ordinated research programmes by the joint FAO/IAEA Division of Nuclear Techniques in Food and Agriculture; IAEA being the International Atomic Energy Agency based in Vienna. In 2012, FAO/IAEA replaced their *Mutation Breeding Manual* with a new book *Plant Mutation Breeding and Biotechnology* edited by Shu et al. (2012). They point out that by the end of 2009 the number of mutant cultivars officially released worldwide had reached 3088 (in 190 plant species), up from a mere 77 in 1964. Many examples can be found in their book. Of the 3088 cultivars, almost 700 were in vegetatively propagated crops and of these, over 90 % were developed using physical mutagens.

Today, as more and more plant genomes are sequenced, mutants are indispensable and powerful tools for determining the function of plant genes. Researchers can link the gene with the phenotype by clarifying the mutant phenotype and the causative gene responsible for that phenotype (Morita et al. 2009). As a result, much more is known about the genetic changes brought about by the various mutagens used. Furthermore, this can inform the choice of mutagen for use in breeding. Consequently mutation breeding is now a more attractive option for the induction of variation for a specific trait in a given crop, particularly in successful cultivars of that crop. The aim is to produce a new cultivar more quickly and

efficiently than could be done through sexual hybridization, particularly when the desired trait would need to be introduced from wild relatives of the crop species. The outcome may be a new cultivar in its own right or a parent for use in hybridization programmes. Two examples will suffice. In soybean, mutation breeding was undertaken in 1987 after an evaluation of cultivars and landraces failed to identify any major genes that would dramatically alter its fatty acid composition (Takagi et al. 2012). The result was 46 mutant lines of interest for use in further hybridizations and germplasm development. In Basmati rice, famous for its quality traits, mutation breeding was an attractive proposition because new cultivars would be classified as traditional Basmati provided they retained the quality characteristics of their parents (Rao et al. 2012). The outcome was the release in India in 2005 of cultivar Geetanjali, a high yielding, early maturity, semi-dwarf, non-lodging Basmati rice with the quality of its parent, cultivar Basmati 370. In contrast it took more than two decades of conventional breeding to combine the high quality traits of traditional basmati with the semi-dwarf, high-yielding attribute that resulted in the world's first high-yielding dwarf basmati cultivar Pusa basmati-1 (Virmani and Ilyas-Ahmed 2007).

Advances in molecular biology coupled with the sequencing of plant genomes are now opening up new possibilities for mutation breeding, including TILLING and targeted genetic alteration, as we shall see later in this chapter. The following account draws heavily on the book edited by Shu et al. (2012) to which the interested reader is referred for considerably more information than can be covered here.

Today mutants are usually produced using radiation, chemicals, T-DNA or transposons. However, it is important to appreciate that mutagenesis is the outcome of the initial direct or indirect interaction of the mutagen with the plant's DNA and any subsequent repair that takes place. The repair mechanism may be error free or may result in heritable change in the genetic material. Before considering the mutagens and the kinds of mutants that they produce, a brief introduction on the design of mutation experiments and breeding programmes is in order, together with a brief mention of mutagenesis of organelle genomes.

## Mutagenesis of Organelle Genomes

In most crop species it is difficult to generate variation in organelle genomes by hybridization because of a lack of natural variability, mono-parental inheritance and reduced or null recombination. Hence mutagenesis is an attractive proposition for a few important traits such as cytoplasmic male sterility (mitochondrial genome) for hybrid seed production and triazine herbicide tolerance (plastid genome). However, mutation induction is more difficult than for nuclear genes and relatively few examples are given in Shu et al. (2012).

## Design of Mutation Experiments and Breeding Programmes

### *Choice of Mutagen and Plant Material*

Decisions are required on the choice of mutagen and the plant material to be mutated; and for physical mutagenesis, whether chronic or acute irradiation is going to be used. Chronic irradiation is the continuous exposure of growing plants over extended periods of time to relatively low doses of irradiation; for example, a rate of 0.25–1.5 Gy/day for 100 days giving a total dose of 25–150 Gy. Acute irradiation is a single exposure at higher doses over a short period of time; for example, a rate of 2.5 Gy/h for 30 h giving a total dose of 75 Gy. Dose rate, then, is simply the radiation dose absorbed per unit of time. Currently most mutation induction is acute in nature. We will see that the choice of mutagen affects the frequencies of different types of mutation and hence the likelihood of loss, gain or alteration of gene function. For example, many breeders have used chemical mutagens (mainly EMS) to create herbicide-tolerant mutants at a frequency higher than the spontaneous rate, whereas to date there has been no commercial herbicide-tolerant crop created and developed through gamma irradiation (Tan and Bowe 2012).

Plants that are produced directly from seeds (or gametes) treated with a mutagen are designated M<sub>1</sub> plants; the untreated seeds being the M<sub>0</sub> generation. Likewise the terms M<sub>0</sub>V<sub>0</sub> and M<sub>1</sub>V<sub>1</sub> (or MV<sub>1</sub>) are used for vegetatively propagated plants, where the treated material can be a wide range of plant propagules or aseptic cultures including cell suspensions, somatic embryos and micropropagation systems. Today an important consideration is whether or not *in vitro* mutagenesis is possible and desirable. This is the process whereby explants or *in vitro* cultures are treated with the mutagen. It can be applied to both seed crops (haploid cultures) and vegetatively propagated crops, but is particularly advantageous for the latter since mutants often originate from single somatic cells, large populations of which can be handled. The success of any *in vitro* mutagenesis programme does however depend upon established and reproducible *in vitro* plant regeneration procedures. Furthermore, to benefit from large populations, efficient *in vitro* screening for desired mutants is required. The mutagenic treatment of isolated microspores at the uni-nucleate stage of development, followed by *in vitro* selection for particular traits, has been developed for use in rapeseed and other *Brassica* species using both physical and chemical mutagens (Szarejko 2012). The latter can be dissolved in the culture medium. Imidazolinone-tolerant rapeseed was developed by treating microspores with a chemical mutagen (ethyl nitrosourea) followed by culturing the microspores in a medium containing an imidazolinone herbicide (Tan and Bowe 2012).

## **Doses and Dose Rates**

Physical and chemical mutagens affect  $M_1$  and  $MV_1$  plants in two ways: physiological damage to cells (primary injury) and DNA damage that can result in mutagenesis. Preliminary assays are required to determine the appropriate doses (and dose rates) for mutagenesis in order to achieve a balance between the extent of primary injury and the frequency of mutations. Balance is also required between the frequency of desired mutations and the mutational load from the induction of other undesired ones; and this often necessitates lower doses. When using physical mutagens these assays are known as radiation sensitivity testing. With chemical mutagenesis it may take a few hours for the mutagen to reach the target meristematic cells (ones capable of cell division) during which time it can be lost by degradation and participation in other chemical reactions. Hence exposure time (usually a few hours) as well as concentration needs to be determined. Mutagen doses that give 30 % lethality ( $LD_{30}$ ) in the treated material (e.g. seeds, seedlings, explants and embryos from microspores) or 30 % average growth reduction ( $RD_{30}$ ), relative to untreated controls, are commonly used to generate mutant populations for crop improvement. Experiments have shown that the effects of mutagens are influenced by both genetic and environmental factors, particularly oxygen and water content.

## ***Dissociation of Chimeric Structure of $M_1$ Plants***

As mentioned in Chap. 3, the desired outcome of mutation breeding is usually a homohistic mutant in which the whole plant has the mutant genotype. This means that the chimeric structure of the  $M_1$  plants obtained by mutagenesis needs to be dissociated. In seed-propagated crops this is usually done by self-pollination of  $M_1$  plants and identification of mutated plants in the  $M_2$  progeny. An alternative for some crops (e.g. wheat, rice and barley) is the use of  $M_1$  plants as donors of gametic cells for doubled haploid production. In vegetatively propagated crops use is made of plant regeneration via somatic embryos or adventitious buds from leaf or root cuttings or from explants which are cultivated *in vitro*. Hence in designing a mutation breeding programme, one needs to consider the population sizes required in both the  $M_1$  ( $M_1V_1$ ) and  $M_2$  ( $M_1V_2$ ) generations for achieving the desired frequency of mutations. First, however, we need to deal with the complication of the number of initial cells.

### ***Number of Initial Cells***

The initial cells (see Chap. 3) of the mature embryo that develop into reproductive organs are called germ track cells, and their number is referred to as the genetically effective cell number. In barley 2–4 initial cells give rise to each of the main spikes which can number 4–5 in total (Gaul 1964). Even with induced mutagenesis, any particular mutation is still a rare event and hence very unlikely to occur simultaneously in both copies of a gene in a diploid cell, so the mutation will be present in the heterozygous condition. If the mutation rate of a gene at a particular locus is  $\mu$ , the probability of mutation of one or both copies of the gene is  $[2\mu(1-\mu)+\mu^2]$  which is approximately  $2\mu$  as  $\mu$  is likely to be about  $10^{-4}$ . In other words, the mutation rate per cell is double the mutation rate per gene. Furthermore, the probability of the simultaneous induction of mutations at two or more initial cells is negligibly small so that any inflorescence that possesses the mutated allele is often in a chimeric state with both mutant and normal tissue sectors.

With a recessive mutant in one initial cell, a 3:1 ratio of normal to mutant seedlings in the spike progeny ( $M_2$  generation) indicates that the reproductive tissue (male and female) came from this one initial cell. However, ratios of 7:1, 11:1 and 15:1 would indicate that the mutated initial cell was one out of two, three and four, respectively (or in general, one mutant out of  $4k$  offspring for  $k$  initial cells). If all of the florets on a spike come from an initial cell of genotype  $Aa$ , where 'A' and 'a' are the normal and mutated alleles, then on selfing:

$Aa \times Aa$  gives 3  $A-$  to 1  $aa$  progeny, where  $A-$  is  $AA$  or  $Aa$ .

However, if the florets have come from two, three or four initials, only one of which was heterozygous ( $Aa$ ) for the mutant, on selfing the florets one would get:

( $AA \times AA$ ) : ( $Aa \times Aa$ ) in the ratios 1:1, 2:1 and 3:1 [ $(k-1):1$ ], and hence spike-progeny ratios of  $A-$  :  $aa$  of 7:1, 11:1 and 15:1 [ $(4k-1):1$ ]. In all situations the ratio of heterozygous to homozygous mutant ( $Aa:aa$ ) is 2:1.

If there were four initials, but only the mutated one survived (the others being killed by the treatment), once again there would be a 3:1 ratio of  $A-$  to  $aa$ . If on the other hand there was diplontic (in vegetatively growing tissue) cell selection and haplontic (at reproductive stage) gamete selection against the mutant, the frequency of  $aa$  genotypes would be reduced. Hence with spike progenies of 20–30 offspring, it can be difficult to infer precisely what has occurred. Nevertheless, the observed frequency of mutants per progeny is called the segregation frequency, in contrast to the segregation ratio expected from Mendelian inheritance.

The above considerations apply to a hermaphrodite species that can be self-pollinated. Some cross-fertilizing crop species such as maize have separate male and female flowers which are derived from different cells in the seed embryo. Hence either the 'male' or 'female' initial cell will be  $Aa$ , but not both, and on being self-pollinated  $AA \times Aa$  will produce  $AA$  and  $Aa$  offspring in the  $M_2$ . Homozygous mutant plants  $aa$  appear in the  $M_3$  generation when  $Aa M_2$  plants are self-pollinated. Hence in maize, pollen mutagenesis can be attractive. Subsequent pollination with

the mutagenized pollen will give an  $M_1$  population of plants comprising 'AA' and 'Aa' individuals which are not chimeric. This means that induced mutations can be discovered in these  $M_1$  plants by TILLING, which is explained in a later section. DNA is extracted from each  $M_1$  plant or from a bulk of 10–12  $M_2$  plants obtained from selfing the  $M_1$ . The  $M_2$  plants are random mated to provide a seed stock at  $M_3$ . Several imidazolinone-tolerant maize mutants (expressed in heterozygotes) were developed by soaking pollen with a chemical mutagen (EMS), followed by pollinating maize plants with the treated pollen and spraying  $M_1$  plants with the herbicide (Tan and Bowe 2012).

In cross-pollinated crops, to obtain mutants, it is essential to adopt a breeding programme that forces inbreeding because recessive mutants would otherwise remain hidden in the heterozygous condition. If the frequency of the mutant gene (' $a$ ') was  $\mu$ , ' $aa$ ' genotypes arising from random mating would have a negligible frequency of  $\mu^2$ . One solution is the crossing-within spike progeny method in which the plants from each spike are grown close together in isolation (for example by bagging) and allowed to cross-pollinate. The  $M_1$ -spike derived from a mutated seed ( $Aa$ ) with one initial cell, will produce 'A' and 'a' eggs in equal frequency, provided they are equally viable. However, these will be pollinated by 'A' pollen from other plants as the frequency of 'a' pollen will be negligibly low. Hence the  $M_1$ -spike progeny will be an equal mixture of 'AA' and 'Aa' plants. Allowing these to intermate will produce offspring ( $M_3$ ) in which the frequency of 'aa' plants is 1/16 (= 0.0625).

In crops that reproduce only by vegetative reproduction, mutagenesis of a heterozygous cultivar will primarily unmask recessive traits by 'knocking-out' wild type dominant alleles at the heterozygous loci. Mutation frequency will be too low to simultaneously affect both alleles at a locus. Dissociation of the chimeric structure of the  $M_1V_1$  plants will occur over vegetative generations. For example, let us consider irradiated microcuttings. The axillary shoots which develop after irradiation are cut and transferred to fresh medium where they grow and produce new axillary buds and shoots, and the process is repeated until the plants are transferred to soil. The original microcuttings will also produce further axillary shoots which can also be propagated. By  $M_1V_4$  (or up to  $V_6$ ) chimeras will have dissociated and plants can be screened for desirable mutants. In contrast, dissociation of chimeras is not required if cell cultures are mutagenized and *in vitro* plant regeneration occurs via somatic embryogenesis and organogenesis from single cells.

### ***Population Sizes in Mutation Programme with Vegetatively Propagated Crops***

With vegetatively propagated crops, the following formula (Suprasanna et al. 2012) is used to determine the number of plant parts ( $X$ ) that need to be treated in order to obtain a certain number of plants ( $P$ ) for field selection:

$$X = P / (\alpha b c d)$$

Where  $\alpha$  is the expected proliferation rate in each sub-culture,  $b$  is the number of sub-cultures,  $c$  is the proportion of cultured plants expected to root and  $d$  is the proportion of rooted plants expected to survive. The number of plant parts then needs to be divided by the mutation rate for the desired trait in order to decide how much material to mutagenize. The frequency range quoted by Suprasanna et al. (2012) is 0.1–1.0 %, so that among 1000 plants from 1–10 individuals would have the desired trait; but this needs to be determined experimentally. Using realistic figures for the other factors in the formula ( $\alpha = 3.3$ ,  $b = 4$ ,  $c = 0.85$  and  $d = 0.90$ ), they suggest that 1000 plants for field selection could be obtained by treating as few as 100 plant parts. Next, however, we come up with much larger numbers for self-pollinated crops; but these are based on mutation rate per cell, not per trait. In other words, one has to be clear about what one is seeking in the way of mutations.

### ***Population Sizes in Mutation Programme with Self-Pollinated Crops***

With self-pollinated crops we need to determine the sizes of  $M_1$  and  $M_2$  populations required to ensure a reasonable probability of identifying a desired mutant. Let

$m$  be the number of  $M_1$  plants and  $M_2$  lines obtained by self-pollination (i.e. no losses) and

$n$  the number of plants from each  $M_1$  spike (a spike progeny). Then

$T_2 = mn$  is the total number of  $M_2$  plants. Let

$p_1 (= 1 - q_1)$  be the mutation rate per cell and

$p_2 (= 1 - q_2)$  the segregation ratio of the mutant trait ( $= \frac{1}{4} = 0.25$  for a recessive trait).

If there are  $k$  initials, where  $k > 1$ , we also need the segregation frequency ( $p_3$ ) in a spike progeny:

$p_3 (= 1 - q_3) = p_2/k$ , assuming mutations in more than one initial are negligible.

Furthermore, the probability of a mutation in a spike arising from  $k$  initials is  $k$  multiplied by the mutation rate per cell ( $p_1$ ), again assuming mutations in more than one initial are negligible:

$kp_1$  is mutation rate per spike.

The probability ( $P$ ) that one or more mutants are detected among the  $M_2$  plants is one minus the probability that no mutant is detected in  $m$  lines ( $\psi^m$ ).

$$P = 1 - \psi^m \text{ or}$$

$$m = \log(1-P)/\log(\psi)$$

No mutant is detected in a line (probability =  $\psi$ ) when there is no mutation in the spike initials ( $1 - kp_1$ ) or the mutation occurs ( $kp_1$ ) but is not detected ( $q_3^n$ ). Hence

$$\psi = (1 - kp_1) + kp_1 q_3^n = 1 - kp_1 (1 - q_3^n)$$

Therefore the total number of  $M_2$  plants required to detect a mutant with probability  $P$  is

$$T_2 = nm = n \log(1-P)/\log(\psi)$$

For given  $P$ , say 0.95 (95 % chance of detecting mutant), the total number of plants required is a function of  $n$  and  $k$ . Although not demonstrated here, it can be shown that the minimum occurs when  $n = 1$ ; in other words, when one grain is taken from each spike (the one-spike-one-grain method).

When  $n = 1$ , we have:

$$T_2 = \log(1-P)/\log(\psi),$$

where

$$\psi = 1 - kp_1(1 - q_3) = 1 - kp_1 p_3 = 1 - kp_1 p_2/k = 1 - p_1 p_2$$

so that  $T_2$  is now independent of  $k$ , the number of initials.

For a recessive trait  $p_2$  is 0.25, so if we choose  $P$  to be 0.95, and if the mutation rate per cell is  $10^{-4}$ , then  $T_2$  is 119,600 plants. Hence this is the minimum number of  $M_2$  plants that have to be raised from 119,600 selfed  $M_1$  plants, one from each plant. There may however be good reasons for raising more than one plant per progeny, such as avoiding loss of a mutant accompanied by severe or complete sterility; sister heterozygous  $M_2$  plants can be selfed to recover the mutant at  $M_3$ .

If for practical reasons it is not feasible to grow a densely planted  $M_1$  generation, it can be argued that the above calculations should be done for the total number of  $M_1$  plus  $M_2$  plants:

$$T_{1+2} = m + nm = (n + 1)\log(1-P)/\log(\psi)$$

When  $n = 1$ , the total number of plants is simply doubled and again the result is independent of  $k$ . However, the calculations in Shu et al. (2012) show that the optimum number of plants per progeny is not one, and becomes larger as the number of initial cells increases. This is another argument in favour of spike-progenies of size greater than one; in fact usually 10–20 in the  $M_1$ -spike progeny method.

What these considerations demonstrate is the importance of mutation rate in determining the practical population size required, and hence the need to find a mutagenic treatment that gives a high mutation rate without too much primary injury to the treated material. Next therefore we will consider the estimation of mutation rate. Later we will see examples of actual population sizes used in the M<sub>1</sub> and M<sub>2</sub> generations of mutation breeding programmes and the methods used for recognizing and hence selecting desired mutations.

### ***Estimating Mutation Rates***

The expected number of M<sub>2</sub> progenies that include one or more mutants is the total number of progenies ( $m$ ) multiplied by the mutation rate per spike ( $kp_1$ ) and also by the probability of one or more mutants being detected ( $1-q_3^n$ ) in a progeny of size  $n$ :

$$mkp_1(1-q_3^n)$$

The expected number of mutants per progeny for progenies in which one or more mutants are detected segregating is the progeny size ( $n$ ) multiplied by the frequency of mutants ( $p_3$ ), but corrected for the progenies where segregation is not detected by chance ( $q_3^n$ ):

$$np_3/(1-q_3^n)$$

Hence the expected total number of mutants ( $T_{mut}$ ) in the M<sub>2</sub> is these quantities multiplied together:

$$T_{mut} = mkp_1np_3 = mnp_1p_2 \text{ (as } p_3 = p_2/k\text{)}$$

From this  $p_1$  is estimated as:

$$p_1 = T_{mut}/(T_2 p_2)$$

If  $p_2$  equals 0.25 (a recessive mutation), then four times the frequency of mutants in the M<sub>2</sub> is an estimate of the mutation rate, as proposed by Gaul (1960). In contrast, it should be noted that the frequency of mutant progenies ( $p_1^*$ ) is not strictly an estimate of the mutation rate. It is simply  $mkp_1(1-q_3^n)$  divided by  $m$ :

$$p_1^* = kp_1(1-q_3^n)$$

Hence if  $k$  is one and  $n$  is small,  $p_1^*$  is smaller than  $p_1$ ; whereas if  $k > 1$  and  $n$  is large,  $p_1^*$  is larger than  $p_1$ .

Today, in the age of DNA sequencing, mutation frequency ( $f$ ) can also be expressed as the number of mutations per 1000 kilobases (kb) of DNA:

$f = (n \times 1000)/(ij)$ , where  $n$  is the total number of DNA lesions detected in  $j$   $M_2$  individuals each assessed for the same DNA fragment of length  $i$  kilobases.

The frequencies found in the TILLING experiments (chemical mutagenesis) summarized later in Table 16.2 range from 1 in 1000 kb for barley up to 26 in 1000 kb for wheat, with values of 3–4 in 1000 kb for rice. This means that for a given gene of length 1 kb, for example, 3–4 mutations will be found among 1000  $M_2$  plants. However, as rice has a diploid genome size of 389 Mb, this also means that each plant may have a total of 1200–1600 mutations (lesions). Even allowing a tenfold reduction for non-coding regions, this probably means mutations in 120–160 protein-coding genes in addition to the target one; and hence additional breeding work to separate the mutant of particular interest from the other ones. Likewise in barley, Caldwell et al. (2004) estimated an average of 5000 mutations per genome. In contrast, lower rates of one mutation per 6190 kb have been reported for a TILLING experiment where seeds of the *japonica* rice cultivar ‘Koshihikari’ were irradiated with 500 Gy of gamma rays for one hour (Sato et al. 2006). Twenty five regions from 1.0 to 1.5 kb in length were screened in the  $M_2$  plants derived from 2130  $M_1$  plants, one from each. Six mutations were identified: four single nucleotide substitutions and one 2-bp and one 4-bp deletion.

### ***Checking for Contamination***

Finally, having completed a mutation breeding programme, it is important to confirm that one really has a mutant and not a contaminant (accidental mixing of seed or outcrossing). This is most easily done by comparing the mutant and the genotype (cultivar or breeding line) that was used for mutagenesis using molecular markers. As induced mutations are relatively rare, a mutant should have an almost identical molecular fingerprint to its parent. Fu et al. (2008), for example, found no differences in rice between seven mutants (induced by gamma-rays) and their parents at 340 microsatellite (SSR) loci. In contrast, even two similar cultivars differed at 5.6 % of these loci. They had expected no differences for true mutants based on an assumption of a weighted spontaneous microsatellite mutation rate of  $3.5 \times 10^{-5}$  per locus per generation and a sixfold increase after mutagenic treatment. In addition, particular characteristics of the parents allowed them to detect contaminants at  $M_2$  from the 0.8 % outcrossing which they estimated had occurred among their  $M_1$  plants.

Sometimes a mutant can become a new cultivar in its own right, on other occasions it is used as a parent in further breeding work. Where different mutations have been obtained in different lines, they can be combined through sexual reproduction. In contrast, with vegetatively propagated crops, further mutations can be induced in a mutant clone through another round of mutagenesis and selection.

## Radiation-Induced Mutants

Radiation acts on living cells by releasing energy along its path and can be divided into two broad categories, ionizing and non-ionizing radiation. The latter refers to any type of electromagnetic radiation that lacks sufficient energy per quantum to ionize atoms or molecules; for example, ultraviolet light is a mutagen, but because of its low penetrability has been confined to the treatment of single cell layers in culture such as isolated microspores. Ionizing radiation consists of highly energetic particles, or electromagnetic radiation, that can eject (ionize) at least one electron from an atom or molecule. Ionizing ability depends on the energy of the individual particles or quanta of radiation. Linear energy transfer (LET) is the energy transferred per unit length of the track. X-rays and gamma ( $\gamma$ )-rays are examples of ionizing electromagnetic radiation that causes mutations and have LET values of 0.2–2 keV/ $\mu$ m. Examples of ionizing particles are alpha ( $\alpha$ ) particles, beta ( $\beta$ ) particles, neutrons, low-energy ion beams and high-energy ion beams. Carbon (C) ion beams have LET values of 23 keV/ $\mu$ m whereas iron (Fe) beams have values of 640 keV/ $\mu$ m. Alpha particles and neutrons also have high LET whereas electron beams (beta particles) have low LET.

Since the 1960s gamma-rays have been the most widely used physical mutagen, although X-irradiation may become the preferred option again because modern X-ray machines have far fewer security and safety issues. About half of all mutant cultivars carry genes mutated by gamma-rays. FAO/IAEA did have a Neutron Seed Irradiation Programme (1966–1974), but the use of fast neutrons as a mutagen has been limited, as has the use of alpha and beta particles. In contrast, since the 1990s there has been increasing interest in heavy-ion beam radiation mutagenesis and ion implantation mutagenesis. The latter technique, which was developed in China, may increase in popularity as it is relatively simple and does not need expensive equipment. Ions are extracted from plasma produced by an electrical discharge in gas or vapour. Twenty two cultivars in total of tomato, maize, wheat, rice and soybean have already been released and grown in China. Details can be found in the chapter by Feng and Yu in *Plant Mutation Breeding and Biotechnology* edited by Shu et al. (2012). Here we are going to consider gamma-ray and heavy-ion beam induced mutants in more detail.

### ***Gamma-ray Mutagenesis***

The commonly used sources of  $\gamma$ -rays are from the disintegration of the radioisotopes cobalt-60 ( $^{60}\text{Co}$ ) and caesium-137 ( $^{137}\text{Cs}$ ). This means that specialist facilities are required for mutagenesis. Gamma cell irradiators in purpose built laboratories are commonly used for a single exposure of seed to radiation. In contrast, prolonged exposure of growing plant material is done from a gamma source placed in a

specially designed phytotron (growth room), or glasshouse, or outdoor field surrounded by a high shielding dyke and containing other safety features.

Gamma-rays produce their effects through both direct damage (interactions) to DNA (bases and sugar-phosphate backbone) and indirect damage via interactions with the surrounding water to produce reactive oxygen species (ROS), primarily hydroxyl radicals. These free radicals can also damage the sugar-phosphate backbone and the DNA bases, with modifications that lead to base substitutions. More than half of the biological effect of low LET ionizing radiation such as  $\gamma$ -rays (and X-rays) is the result of this indirect action because of the abundance of water. The mutagenic outcomes result from the repair done by various mechanisms, which are activated to deal with the single and double strand breaks in the DNA and the changes to the DNA bases. Hence different types of mutations might be expected and are indeed found. Nevertheless, where each mutation is the result of a single ionization (single hit) there is a linear relationship between dose of radiation and yield of mutations, whereas when two events (chromosome breaks) are required for chromosomal rearrangements, they are produced in proportion to the square of the dose. They are also more likely if the dose is given at high intensity for a short duration, rather than low intensity for a long duration, so that the second break occurs before the first is repaired.

### ***Gamma-ray Mutants***

Morita et al. (2009) made a detailed study of 24 gamma irradiation-induced mutations in six cultivars of japonica rice. Seventeen resulted from irradiation of seed at doses from 200 to 300 Gy and rates from 20 to 30 Gy/h, three from flowers (100 Gy at 2 Gy/h) and four from plants (200–320 Gy at 10–16 Gy/day). The phenotypes of the mutants are shown in Table 16.1. Mutation analysis was done by PCR performed on the causative genes using primers amplifying approximately 1 kbp, followed by sequencing. The technique was modified when the whole or part of the coding region did not amplify. Fifteen mutations (62.5 %) were small deletions (1–16 bp), four (16.7 %) were large deletions (9.4–129.7 kbp), three (12.5 %) were base substitutions and two (8.3 %) were inversions (Table 16.1). As base substitutions can produce silent mutations, the proportion of base substitutions may have been higher than reported. The base substitutions were transversions, but transitions have been found in other studies. G:C to T:A transversions can be explained through the action of the reactive oxygen species generated by gamma rays, and T:A to A:T transversions may occur through a similar mechanism. Only one mutant, a large deletion, also had an inserted base at the rejoin which suggests that rice, like *Arabidopsis* (but unlike tobacco), does not often use double-strand break (DSB) repair mechanisms involving ‘filler DNA’ insertions (short scrambled sequences new to plant). Microhomology was found in 14 of the 19 (73.7 %) deletion events, and in three of the four rejoining sites of the two inversions. This suggests that the repair of gamma irradiation-induced DSBs

**Table 16.1** Gamma ray mutants in rice (Morita et al. 2009) (*GA* gibberellins)

Mutant	Number	Deletions	Base substitutions
Waxy	6	2 and 5 bp (frameshift)	GT to GA (at splice donor site) transversion
		6 bp (Leu–Leu–Cys to Arg)	
		9.4 kbp (whole gene)	
		1284.8 kbp inversion with 1 and 4 bp deletions at joins	
Short plastochron	4	5 and 5 bp (frameshift) 3208.5 kbp inversion with 2 and 75 bp deletions at joins	GAG to GTG (glutamic acid to valine) transversion
GA-insensitive dwarf	2	1 bp (frameshift) 42.2 kbp (whole gene)	
GA-sensitive dwarf	5	1–16 bp (frameshifts)	
Chlorophyll b deficiency	2	1 and 3 bp	
Glutelin deficiency	4	1 and 1 bp (frameshift?) 129.7 kbp (whole gene?) plus filler DNA	GAA to TAA (Glu to stop) transversion
$\alpha$ -Globulin deficiency	1	62.8 kbp (whole gene?)	

frequently uses microhomology (over a few bases). The small deletions were typical of those found in other studies with low LET radiation and are in contrast to the longer deletions of 300 bp to 12 kbp found with high LET radiation such as neutrons, which induce DSBs at closer intervals than gamma rays. The small deletions were probably the result of a single DSB being repaired using endonucleases. The large deletions and inversions were probably the result of two DSBs occurring simultaneously on the same chromosome, which were rejoined by NHEJ, with either loss or inversion of the segment between the breaks. It is assumed that very large deletions cause lethal effects in the gametophyte and hence are not observed in the next generation.

### ***Examples of Breeding Programmes Using Gamma-ray Mutagenesis***

The Basmati rice programme mentioned earlier provides a practical example of mutation breeding using gamma-ray mutagenesis (Rao et al. 2012). Batches of between 150 and 200 g of seed of three cultivars were treated with three doses of gamma-rays (100, 150 and 200 Gy) to secure around 7000–8000 plants per treatment in the M<sub>1</sub> generation. Around 100,000 plants per cultivar (300,000 in total) were raised from bulked seed in the M<sub>2</sub> generation. Visual selection for early flowering plants (plants were tagged) that were non-lodging and had a good panicle

type resulted in a total of 16,860 M<sub>3</sub> families. From the M<sub>3</sub> generation onwards, panicles from four or five single plants per selected progeny were grown separately in the next generation. Selection over generations took place for grain quality, non-lodging and semi-dwarf stature, and after M<sub>7</sub> just 20 lines remained; 3, 7 and 10 from each of the three cultivars. Observational and replicated yield trials took place at the Central Rice Research Institute in India in the wet and dry seasons, before 3 years of mandatory multi-lokalional trials in North West India. Mutant line CRM 2007-1 was released as cultivar Geetanjali in 2005. Hence the size (300,000 M<sub>2</sub>), timescale (12 generations) and success rate (1 in 300,000) of the mutation breeding programme was similar to that of a pedigree inbreeding programme for rice. In contrast, a lower dose (33.75 Gy) and smaller population size (2000 dormant buds on scions for grafting) was used for mutation breeding of sweet cherry in Turkey in 2001 (Kunter et al. 2012). A mutation frequency of 4 % observed abnormalities was achieved. By 2008, after 5 years of juvenile growth (considered five mutant generations) and flowering, eight mutants of interest had been identified for further propagation from among the 371 trees in the plantation which had survived from the 2000 grafted buds. These two examples provide a contrast between mutation breeding in a seed and a vegetatively propagated crop.

### ***Heavy-Ion Beam Mutagenesis***

Particle accelerators such as the ring cyclotron were developed in Japan in the 1970s for research in nuclear physics. Subsequently they were used for applied research, including mutagenesis in plant breeding since 1993 (Ryuto et al. 2008). The RIKEN (Wako Institute, Saitama) ring cyclotron can produce ions of carbon, nitrogen, neon, argon, iron and other elements. Atomic ions (atoms deprived of electrons) from radioisotopes are accelerated in a spiral to half the speed of light. Heavy-ion beams have a lot of energy per particle, up to 1000 times more than those of X-rays and gamma-rays, and show high linear energy transfer (LET). Biological samples in containers are placed into cassettes which are automatically moved to the beam position. The treatment given is short, only seconds or a few minutes. Samples can be dry seeds, imbibed seeds, tissue culture material and cultured plantlets.

Heavy-ion beam irradiation is considered a good tool for mutation breeding because at relatively low doses it induces mutations at a high rate without severely inhibiting growth. This is because high-LET radiation causes more localized, dense ionization than low-LET radiation and hence mainly direct damage to DNA. It therefore induces DNA double-strand breaks more effectively than other mutagenic techniques with the resulting mutations determined by the DNA repair mechanism, usually NHEJ. Furthermore, as only one site on the DNA double-strand is broken by the passage of the beam, a recessive mutant lacking just a single gene is produced with high probability. Hence it is easier to alter a single characteristic with heavy ion-beams than with X-rays or gamma rays, and thus obtain a new

cultivar by selecting a mutant with a modification to the target characteristic while retaining the existing valuable characteristics of the unmutated cultivar. Facilities for heavy-ion beam irradiation are now available in China and Italy as well as Japan.

### **Heavy-Ion Beam Mutants**

The first mutant cultivar to be released in 2002, in Japan, was the dahlia "World", with large vivid red flowers compared with the smaller pink ones of the parent cultivar "Miharu". Since then, over 30 Japanese cultivars have been released. In a review of this work, Tanaka et al. (2010) mention new cultivars of ornamental flowers and trees and the isolation of salt-tolerant and low N-fertilizer-requiring mutants of rice. The ornamental flowers were carnation, chrysanthemum, osteospermum, petunia, pink, trenia and verbena, and the new cultivars had novel flower colours and shapes. The new cultivars of trees were *Ficus thunbergii* (the fig genus) with a high capability to assimilate atmospheric nitrogen dioxide and a cherry blossom tree that blooms in all four seasons. Where seeds were irradiated, often 3000–7000 would be treated to give an M<sub>2</sub> population of 30,000–35,000 plants in which to seek mutants.

Tanaka et al. (2010) also discuss the experiment of Shikazono et al. (2005) in which 29 carbon-ion-induced mutants in two genes of *Arabidopsis* were analysed by PCR and by DNA sequencing: 14 had intragenic point-like mutations (11 were 1–100 bp deletions, one was a 1 bp insertion and two had base substitutions) and 15 had intergenic rearrangements such as deletions (six in the range of 5–230 kbp), inversions, translocations and insertions. Of the 15 breakpoints analysed, 11 had 1–29 bp deletions and four had duplications of 1–5 bp. Most of the break and rejoined sites had microhomologies, suggesting that non-homologous end-joining (NHEJ) played an important role in the rejoining of double-strand breaks or clustered damage induced by carbon ions. Tanaka et al. (2010) concluded that chemical mutagens and low LET ionizing radiation will predominantly induce many small DNA modifications, resulting in several point-like mutations on the genome. In contrast, high LET ionizing radiation such as heavy-ion beams will efficiently cause fewer but larger and irreparable modifications locally, resulting in a limited number of null mutations.

Kazama et al. (2011) studied heavy-ion mutagenesis by irradiating dry *Arabidopsis thaliana* (ecotype Columbia) seeds with carbon ions. Their LET values were 22.5 and 30 keV μm<sup>-1</sup> at doses of 250 and 450 Gy with the former and 400 Gy with the latter. With a LET value of 30 keV μm<sup>-1</sup>, they obtained a mutation frequency similar to that achieved with the chemical mutagen EMS, and two and a half times that achieved with X-rays. Analysis of their 22 mutations in five genes revealed 3 transversions, 1 transition, 13 small deletions (<100 bp), 1 small insertion and 4 chromosome rearrangements. They assumed that these were a consequence of the way the double-strand breaks were repaired. There was one

missense and 17 nonsense mutations, plus the four chromosome rearrangements, making a total of 21 null mutations out of 22 mutants. Kazama et al. (2011) concluded that heavy-ion mutagenesis is most useful for producing null mutants whereas the chemical mutagen EMS is best for producing an allelic series with the possibility of finding an allele with a new function. In other words, EMS produces base substitutions that result in missense mutations (change to codon for another amino acid) as well as nonsense (change to stop codon) and silent mutations (change to codon for same amino acid which does not generate any modification in phenotype). If the mutation occurs at the exon-intron interface, modifications to splicing can occur.

## Chemical Mutagenesis

Although a large number of chemical compounds are now known to be mutagenic, alkylating agents were used to produce over 80 % of the plant cultivars in the IAEA database obtained via chemical mutagenesis, and just three of them accounted for 64 % of the cultivars. The three alkylating agents are: ethyl methane sulphonate (EMS), 1-methyl-1-nitrosourea (MNU) and 1-ethyl-1-nitrosourea (ENU). We will therefore consider them in detail before briefly mentioning sodium azide and other chemical mutagens. Chemical mutagens such as alkylating agents are very toxic substances and are also carcinogens. They need to be handled in a safe manner and likewise disposed of in an appropriate manner by fully trained staff.

### *Alkylating Agents*

Seven nitrogen atoms and four oxygen atoms in the four bases of DNA, and an oxygen atom in the phosphodiester backbone, have been identified as the 12 most common targets for the alkylating agents. The mutagenic outcome depends on the extent and mechanism of DNA repair. Alkylation of the N<sup>7</sup> position of guanine is common but the altered base is apparently non-mutagenic. In contrast, when not repaired, alkylated O<sup>6</sup>guanine tends to produce G:C to A:T transitions because O<sup>6</sup>-alkylguanine mispairs with thymine during DNA replication; O<sup>4</sup>alkylthymine can mispair with guanine and lead to A:T to G:C transitions; alkylated N<sup>3</sup>adenine can result in A:T to T:A transversions; and alkylated N<sup>3</sup>cytosine can give rise to C:G to T:A transitions and C:G to A:T and C:G to G:C transversions. In practice, MNU produces almost 100 % G:C to A:T transitions; ENU produces 72 % of these transitions plus 21 % A:T to G:C transitions and 6 % transversions; and EMS produces 93 % G:C to A:T transitions, 1 % A:T to G:C transitions, 2 % transversions and 4 % other mutations (Shu et al. 2012). However, chemical mutagenesis can also result in deletions and insertions and other chromosomal rearrangements resulting from double-strand breaks in the DNA. Hence the total

spectrum of mutations can be wider than one might expect from an initial consideration of the effects of alkylation.

Although a wide range of plant material can be subjected to chemical mutagenesis, seeds are most commonly used, the exception being maize where use of pollen is attractive (as explained earlier). Seeds are sometimes pre-soaked in water or buffer to allow the mutagen to diffuse more rapidly to the target meristems. Examples of mutagenic treatments, including plant material, concentration of mutagen and exposure time can be found in Shu et al. (2012). Typical concentrations are 10–100 mM for EMS, 5–6 mM for ENU and 0.2–1.0 mM for MNU; and typical exposure times are from 1 to 12 h.

### **Sodium Azide**

Sodium azide (Az) was used for the first time as a mutagen by Nilan et al. (1973) in barley, and provides an interesting contrast to the alkylating agents. It is metabolized *in vivo* to a powerful chemical mutagen in crop species such as barley, bean, maize, pea, rice, sorghum, soybean, sugarcane and wheat, but not in the model plant species *Arabidopsis thaliana*. Its mutagenic effect depends on the pH of the treatment solution, being ineffective in alkaline conditions. Hence a pH of 3 is often used for mutagenesis. The mutagen is the amino acid analogue L-azidoalanine. It induces chromosomal aberrations at a very low level compared with other mutagens and has been combined with MNU to produce a very high level of point mutations (e.g. 1.5 mM Az followed by 0.7 mM MNU). The protocol can be found in Shu et al. (2012) along with many examples of its use. Olsen et al. (1993) demonstrated that sodium azide mutagenesis of the barley *Ant18* gene (encodes dihydroflavonol-4-reductase) preferentially generated A:T to G:C transitions.

### **Other Types of Chemical Mutagen**

Nitrous acid reacts with amino groups to form hydroxyl groups and hence deamination of bases: guanine to xanthine and oxanine, adenine to hypoxanthine and cytosine to uracil. Hypoxanthine pairs with cytosine to give A:T to G:C transitions while uracil pairs with adenine to give C:G to T:A transitions. Xanthine and oxanine can pair with cytosine or with thymine, the latter giving G:C to A:T transitions. Inter-strand cross-links are also formed preferentially in 5' CG sequences where the two guanines on opposite strands covalently link through a shared exocyclic amino group. This cross-linking could give rise to various mutations through error-prone excision repair. However, nitrous acid has not been used to any extent in plant genetic improvement.

Base analogues mimic a normal DNA base (5-bromouracil) or base-pair (ethidium bromide) so as to interfere with replication. Thus incorporation of 5-bromouracil in place of thymine induces G:C to A:T and A:T to G:C transitions, and A:T to T:A and C:G to A:T transversions, as well as small indels that cause frameshift mutations. Intercalating agents such as acridinium salts and ethidium bromide can also induce frameshift mutations as they mimic a base-pair and get incorporated into the DNA. Again none of these mutagens have been used to any extent in plant genetic improvement. In contrast, the antibiotics mitomycin C and streptomycin have been used to induce male sterile mutants in a number of plant species, including sunflower (*Helianthus annus*) (Jan and Vick 2006).

## Identification of Mutants

The largest populations that can be screened are those that involve *in vitro* culture. This is because the culture medium can be subjected to appropriate chemical (in the medium) and physical (surrounding environment) stresses that allow only those cells to grow that possess the required tolerance or resistance. Examples of agricultural traits that can be selected in this way are disease resistance (if a toxin is involved), herbicide tolerance, salt tolerance, metal tolerance, cold tolerance, drought (through high osmoticum in medium), enhanced amino acid accumulation (high concentration in medium) and flooding tolerance (anaerobic culture). Clearly not all traits can be selected at the cellular level, and those that can need to be confirmed in appropriate field trials. *In vitro* mutagenesis and selection is best developed in rapeseed and other *Brassica* species where efficient microspore culture protocols are available (Szarejko 2012). A double mutant line for herbicide tolerance (to imidazolinone), produced by crossing two single mutants, has been commercialized in Canada under the name SmartCanola. Interestingly in microspore-derived haploid embryos, fatty acid biosynthesis and accumulation is similar to that in zygote embryos in seeds. Thus one cotyledon of a late cotyledonary microspore embryo can be used for a non-destructive test to evaluate fatty acid composition, whilst the rest of the embryo can be regenerated into a plant. Selection can be practised for increased oleic acid, and an accompanying reduction of linoleic acid, or any other desired fatty acid profile. Regenerated plants can be transferred to soil and once established, treated with colchicine to produce doubled haploids which can then be maintained by self-pollination.

Large mutagenized plant populations are most easily screened for phenotypically expressed mutations such as changes in plant morphology (e.g. height), maturity date, disease resistance and herbicide resistance. Forster et al. (2012) describe, with illustrations, how 21,000 M<sub>3</sub> (and M<sub>4</sub>) families from barley cultivar Optic were screened in the glasshouse and field for a wide range of mutants affecting the following traits: caryopsis, seedling development, vegetative growth (i.e. phytomer development), reproductive growth, inflorescence formation, spike development, meiosis and flowering, and adult plant ripening. Among the 21,000

families, 7204 exhibited at least one, and 4324 at least two, visible mutant phenotypes. EMS at two dosage rates, 20 and 30 mM, had been used for the mutagenesis. Large mutagenized populations have also been screened for changes in plant biochemistry, such as fatty acid composition in oil crops. Today, however, reverse genetics methods such as TILLING are also available.

## TILLING

Large mutagenized populations can be screened for mutational changes in specific genomic sequences using a procedure known as TILLING (Targeting Induced Local Lesions IN Genomes). This is a reverse genetics method in which chemical mutagenesis is coupled with a high-throughput screen of the mutagenized population to detect point mutations (single-base changes) in target DNA sequences. These are usually one or more genes of known sequence which are chosen because of their putative involvement in a particular biochemical pathway or biological process. Single-copy genes are preferred as point mutations are more likely to generate detectable phenotypic alterations. With large genes it is necessary to focus on the stretch of sequence likely to be of most interest as 1.5 kb is the upper limit of the technique. Chemical mutagens are chosen which induce a high frequency of G:C to A:T transitions with the expectation of generating around 50 % missense mutations. In other words, alleles with altered protein function are generated, some of which may be desirable for crop improvement. Hence the technique has practical use in plant breeding in addition to its role in functional genomics where the aim is to deduce function from the phenotypes of induced mutations. The method is most appropriate for diploid species that can be self-pollinated, but can be used on polyploids, particularly allopolyploids such as wheat (Uauy et al. 2009). As chemical mutagenesis has already been dealt with, we will now concentrate on how the mutagenized M<sub>2</sub> population is screened and the results of some TILLING experiments.

### *Screening for Mutants*

DNA is extracted from leaf tissue of each M<sub>2</sub> plant, which is also self-pollinated to maintain each one as an M<sub>3</sub> family. A key aspect of the efficiency of the method is the pooling of aliquots of the DNA samples. These are arrayed into a multi-well microtitre plate containing 96 or 384 wells so that the whole process can be automated. Typically, eightfold pooling is done in one or two dimensions.

Individual samples are arrayed in 8 × 8 grids of 64 wells, up to the population size. The eight rows of the first grid are labelled 1A, 1B... 1H, the eight rows of the second grid 2A to 2H, and so on. Then in one dimensional (1D) pooling, all eight samples in a row are pooled together. Using an eight-channel pipettor, all

## (a) One dimensional plate

Grid number

1r 2r 3r 4r 5r 6r 7r 8r 9r 10r 11r 12r

G	A
r	B
i	C
d	D
	E
r	F
	G
	H

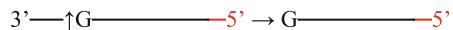
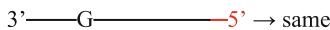
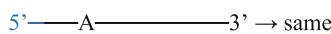
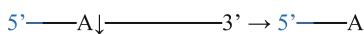
## (b) Two dimensional plate

Grid number

1r 1c 2r 2c 3r 3c 4r 4c 5r 5c 6r 6c

G	A
r	B
i	C
d	D
	E
r	F
or	G
c	H

## (c) Cleavage products on denaturing and separating (blue and red fragments are visualized)



Heteroduplex (mismatch cleaved 3' side)

Heteroduplex (mismatch cleaved 3' side)

**Fig. 16.1** TILLING (*r* row, *c* column)

64 samples from an  $8 \times 8$  plate can be pooled together with eight pipetting steps. They now occupy the first column (Fig. 16.1a) of the 1D 96-well plate ( $8 \times 12$ ). Samples from the second  $8 \times 8$  plate are deposited into the second column, and so on until all 12 columns are filled. Thus the 1D plate contains DNA aliquots from 768 M<sub>2</sub> plants and plates are filled until the whole population is represented.

In two dimensional (2D) pooling, all eight samples in a row are pooled together as before and deposited into the first column of what will become a 2D plate. Next, however, the columns (also A to H) of the first grid are pooled in the same way and deposited into the second column of the 2D plate (Fig. 16.1b). Row and column samples from the second  $8 \times 8$  plate are deposited into the third and fourth columns, and so on until all 12 columns are filled. The 2D plate contains DNA aliquots from

384 M<sub>2</sub> plants and plates are filled until the whole population is represented. Hence the number of DNA pools is doubled but each plant is uniquely represented by a pair of locations in a 96-well plate, thus making unnecessary the analysis of the single components of the pool. Another advantage of 2D pooling is that false-positive errors are reduced because when a mutation is discovered, it will appear in two separate gel lanes.

Screening for mutations then begins with PCR (polymerase chain reaction) amplification of the chosen DNA sequence (usually about 1 kb of amplicon) from the pooled DNAs using fluorescently labelled PCR primers, with one dye for the forward primer and another for the reverse primer. Infrared dyes that are detected at about 700 nm and 800 nm are called IRDye 700 and IRDye 800, respectively.

Following amplification, the PCR products are denatured by heating and allowed to re-anneal. If a mutation is present in the pooled DNA, heteroduplex DNA will form between mutant and wild type DNA strands, as well as homoduplex DNA. Samples are then incubated with a single-strand-specific nuclease (cleaves 3' of mismatch) from celery juice extract (CJE) which breaks one strand of the heteroduplex DNA. The products are denatured and separated by size using denaturing polyacrylamide slab gel electrophoresis, and then visualized on the gel; for example, using a LI-COR DNA analyser system. Mutations are detected by the presence of cleaved bands in ‘different colours’ whose molecular weights sum to that of the full length PCR product (Fig. 16.1c). The cleavage point provides a good guide to the site of mutagenesis. If necessary, pools are then deconvoluted to identify mutant individuals whose target DNA is then sequenced to determine the point mutation.

In allopolyploids such as wheat most genes are represented by two (in tetraploid) or three (in hexaploid) homoeologous copies that share approximately 93–96 % sequence identity (Uauy et al. 2009). Gene duplication limits the use of forward genetics phenotypic screens as the effect of single-gene knockouts are frequently masked by the functional redundancy of homoeologous genes present in the other wheat genomes. However, in TILLING, primers can be designed that are specific to homoeologous copies. Nevertheless, a high probability of identifying truncation mutants is very important because it is generally necessary to cross single mutants in the A and B genome homoeologues to obtain a functional knockout in tetraploid wheat or create the triple A/B/D mutant in hexaploid wheat. Employing missense mutations in these lengthy genetic schemes is risky because if one of the mutations is not effective, it may be sufficient to limit the effect of the combined mutations on function.

Finally, a phenotypic analysis is required to verify the effect of the mutation. This can be complicated by the fact that chemical mutagenesis introduces a large number of point mutations per genome, in addition to those detected at the target locus. Hence the mutants often need to be backcrossed to the genotype that was chosen for the mutagenesis, or outcrossed, to separate their effects from those of other mutations.

## Results

Results from a number of crop species have confirmed that a high frequency of G:C to A:T transition mutations can be induced and detected in targeted DNA sequences and that approximately 50 % of these are missense mutations (Table 16.2). There will be similar numbers of silent mutations and a few truncations as a result of mutation to nonsense codons or to splice junction losses. Among the new alleles created by missense mutations there is the prospect of finding a few with desirable phenotypes for crop improvement. Although point mutations are expected, some major chromosomal aberrations could also be induced, but are likely to be eliminated from the population through their adverse effects on germination, viability and fertility. The extensive mutagenesis in *Arabidopsis* (Greene et al. 2003) also confirmed a 2:1 ratio of heterozygotes to homozygotes for mutants at M<sub>2</sub>, as might be expected from selfing M<sub>1</sub> plants. Very few cells make up the apical meristem in *Arabidopsis* so that almost every M<sub>2</sub> zygote should be from a single lineage: in other words, chimeric M<sub>1</sub> flowers are very rare. Results in wheat (Uauy et al. 2009)

**Table 16.2** Results of TILLING experiments

Crop/Cultivar	Tissue/ Mutagen	Mutations and density per kb	G:C to A:T %	Truncation %/ Missense%/ Silent%
<b>Arabidopsis (Greene et al. 2003)</b>	Seed	1890 mutations from 192 genes in 3072 plants		
Columbia	EMS	1/295 kb	99	5/50/45
<b>Barley (Caldwell et al. 2004)</b>	Seed	10 mutations from 2 genes in 9216 plants		
Cultivar optic	EMS	1/1000 kb	80	0/60/40
<b>Maize (Till et al. 2004)</b>	Pollen	17 mutations from 11 genes in 750 plants		
B73	EMS	1/500 kb	100	0/59/41
<b>Rice (Till et al. 2007)</b>	Seed	57 mutations from 10 genes in 1536 plants		
Cultivar Nipponbare	EMS	1/294 kb	70	0/67/33
Cultivar Nipponbare	Az-MNU	1/265 kb	67	3/63/33
<b>Soybean (Cooper et al. 2008)</b>	Seed	116 mutations from 7 genes in 2833 plants		
Cultivar Forrest	EMS	1/140 kb	92	0/66/34
Cultivar Williams82	EMS	1/550 kb	75	8/33/58
Cultivar Williams82	EMS	1/250 kb	92	4/44/51
Cultivar Williams82	NMU	1/140 kb	90	4/44/51
<b>Wheat (Uauy et al. 2009)</b>	Seed	275 mutations from 11 genes in 2904 plants		
Tetraploid (pasta)	EMS	1/51 kb	99	5/28/67
Hexaploid (common)	EMS	1/38 kb	99	4/40/56

suggest that TILLING is feasible despite wheat's allopolyploid nature, but that it may be easier to find functional knockouts than new alleles for use in crop improvement.

One of the consequences of mutation breeding using chemical mutagenesis is that the background genome contains high levels of random unwanted mutations. These can be eliminated through a backcross process while retaining the mutations of interest. The recurrent parent could be an unmutagenized wild type parent line or a different genetic background of interest for crop improvement. The TILLING programme will have allowed the mutations of interest to be converted into molecular markers (i.e. allele specific markers).

## Causal Mutation for a Given Phenotype

Abe et al. (2012) have developed a method of mapping and isolating mutants called MutMap which is based on whole-genome resequencing of pooled DNA from a segregating population of plants that show a useful phenotype. A mutant is crossed directly to the original wild-type line that was mutagenized, and then selfed, allowing unequivocal segregation in the second filial generation ( $F_2$ ) progeny of phenotypic differences, including those of small effect. For proof of concept, they applied MutMap to mutants of Japanese rice cultivar Hitomebore and identified the genomic positions of two mutations causing pale green leaves, four causing semidwarfism, and one causing male sterility. One mutant line contained two independent mutations in unlinked loci, one for pale green leaves and the other for semidwarfism. It is worth looking in detail at the results for one of their mutants.

The mutants were identified by their phenotypes from a collection of 12,000 EMS-mutagenized lines ( $M_3$ – $M_4$ ) of cultivar Hitomebore. One mutant had a pale-green leaf phenotype with a slightly lower chlorophyll concentration than the wild type (Hit1917-p11). It was crossed to Hitomebore and the  $F_1$  self-pollinated to produce an  $F_2$  progeny in excess of 200 individuals (>100 is considered adequate). Field-grown plants segregated for wild-type and mutant phenotype in the 3:1 ratio expected for a recessive mutation at a single locus. DNA was isolated from 20  $F_2$  progeny showing the mutant phenotype and bulked with equal quantities of each sample. The bulked DNA was subjected to whole-genome sequencing using an Illumina GAIIX sequencer, and resulted in 70 million sequence reads (75 bp) corresponding to >5 Gb of total read length with >12× coverage of the rice genome (370 Mb). The reads were aligned to a reference sequence of Hitomebore using MAQ software6. Mutant Hit1917-p11 contained 1001 transition-type (G → A and C → T) SNPs with high-quality scores, presumably caused by the EMS mutagenesis. This mutation rate of 1 in 370 kb is comparable to the figures in Table 16.2. An index was calculated for each SNP as the ratio of the number of reads with a mutant SNP to the total number of reads corresponding to the SNP; with an expected value of 1 near the causal gene and 0.5 for unlinked loci. When the indices were plotted for all 12 chromosomes of rice, with one exception, for all

parts of the genome they were distributed randomly around 0.5: a larger progeny size for the F<sub>2</sub> bulk would give a smaller variance around this mean whereas increasing the coverage beyond 12× has less effect, as explained by the authors. The exception was a single unique genomic region on chromosome 10 with a cluster of seven SNPs with SNP index of 1. This was the putative position of the causal mutation responsible for the mutant phenotype, as the probability of such a cluster arising by chance is extremely small (the authors calculate less than 1 in  $2.3 \times 10^{-9}$  for a cluster of more than four). One of the seven SNPs with an index of 1 in Hit1917-pl1 (SNP-22981826) corresponded to an exon of the gene encoding chlorophyllide *a* oxygenase (*OsCAO1*) leading to a L253F mutation (codon CTT → TTT; leucine to phenylalanine at position 253 in protein). In a study of a T-DNA insertion knockout of the *OsCAO1* gene, the resulting mutant had lower chlorophyll than wild type, similar to mutant Hit1917-pl1. When the Hit1917-pl1 mutant was transformed with the wild-type *OsCAO1* gene driven by its native promoter in a complementation test, both mutant and wild-type alleles of *OsCAO1* were expressed and a wild-type phenotype was restored. In contrast, wild-type plants transformed with an RNA-interference (RNAi) construct, which silenced the *OsCAO1* gene, showed lower *OsCAO1* transcripts and a paler green phenotype than Hit1917-pl1. These results demonstrated that the Hit1917-pl1 phenotype is caused by the mutation SNP-22981826 identified by MutMap.

## Insertional Mutagenesis

Insertional mutagenesis is a rapid method for mutating genes that can later be easily identified based on knowledge of the flanking sequence tags (FSTs). Jung et al. (2008) reviewed the experimental approaches that have been used to develop rice lines in which genes are randomly tagged by DNA insertion elements. The resulting mutant populations include gene knockout and gene overexpression lines that are useful for determining gene function based on phenotypes; and some lines may be useful to rice breeders. DNA elements that can insert randomly within chromosomes to disrupt gene function include the T-DNA of *A. tumefaciens*, heterologous transposons (Ds and dSpm) and the Tos17 retrotransposon. The insertion of DNA elements into coding regions often leads to complete loss of gene function. If the gene carries out an essential function then the mutation will be lethal, preventing subsequent phenotypic studies. Another type of insertion population consists of ‘activation tagged’ lines that carry gain-of-function phenotypes, produced by insertion of gene cassettes containing a strong enhancer element near one end. Another gain-of-function approach called the ‘FLcDNA over-expresser (FOX) gene-hunting system’ has recently been developed. The combined total of the five mutant populations in 2008 was approximately 500,000 lines, with one to ten copies of mutations per line. If the insertions were evenly distributed across the genome, this FST collection would be predicted to target insertions in 99 % of rice loci. Further explanations of the methods can be found in the review by Jung

et al. (2008), and an update on progress in the review by Wang et al. (2013). Here we will simply mention the strategy for T-DNA insertional mutagenesis.

Embryonic calli are co-cultivated with *Agrobacterium tumefaciens* carrying a T-DNA vector for random insertional mutagenesis (see next chapter for details). After picking transgenic lines using selectable markers, they are regenerated and transplanted to a greenhouse where they flower and set seed by self-pollination. Because the sequence of the inserted element is known, the gene in which the insertion has occurred can be recovered using various cloning or PCR-based strategies. Extraction of DNA from pooled leaves from first generation progeny can be used for the isolation of flanking genomic sequences of inserted T-DNA. The phenotype of the T-DNA insertional mutant allele can then be characterized. By 2008, 172,500 FSTs had been generated worldwide and were estimated to cover 27,551 (48 %) out of the predicted 57,142 genetic loci in rice (Jung et al. 2008).

## Gene Silencing

In the next chapter we will see that transgene-mediated gene silencing through RNA interference (RNAi) offers a directed way of achieving some desirable phenotypes for crop improvement by inactivating one or several specific genes. Such gene silencing mimics recessive loss of function mutations, but in a dominant way so that it can easily be used in polyploids such as potato, in  $F_1$  hybrid cultivars, and also for silencing whole families of genes. Several related approaches act through small silencing RNAs (sRNAs) derived from double-stranded RNA precursors. One example is artificial micro-RNA (amiRNA) technology that exploits endogenous miRNA precursors to preferentially generate a single specific sRNA *in vivo* which induces cleavage of the target mRNA opposite positions 10 and 11 of the miRNA. Warthmann et al. (2008) have demonstrated successful gene silencing by amiRNAs in an indica and a japonica cultivar of rice. Lusser et al. (2012) reported that the introduction to plant cells of genes encoding RNAs homologous to promoter regions, results in the formation of small double-stranded RNAs that induce methylation and silencing of the homologous sequences; in other words, gene expression is modified epigenetically.

## Site-Directed DNA Sequence Modification (Genome Engineering)

The methods of mutagenesis considered so far in this chapter have one feature in common, for all practical purposes they can be considered to occur at random in the genome. Site-directed DNA sequence modification is a major step forward that is becoming feasible in crop species. A reference genome sequence, efficient genetic

transformation, and good functional genomics resources are all desirable. Engineered nucleases are designed to catalyze a double-stranded break (DSB) at a specific location in the genome, and this stimulates the desired DNA modifications to occur at or near the break site. The following brief account is taken from the reviews by Curtin et al. (2012) and Lusser et al. (2012), and the EFSA scientific opinion addressing the safety assessment of plants developed using Zinc Finger Nuclease 3 and other Site-Directed Nucleases with similar function (EFSA 2012b).

## ***Engineered Nucleases***

Zinc Finger Nucleases (ZFNs), transcription activator like effector nucleases (TALENs) and meganucleases (MNs) can all be regarded as types of site-directed nucleases (SDNs) in terms of their modes of action. Meganucleases are also called LAGLIDADG homing endonucleases (LHEs). In all SDNs one domain of the protein is responsible for binding to a specific DNA-sequence while a second domain is responsible for the DNA-cleavage at the binding site. ZFNs and TALENs are proteins which have been engineered to combine two such domains whereas metanucleases are naturally occurring proteins with both domains, but the latter has proved difficult to engineer to target novel DNA sites. Both ZFNs and TALENs are designed to work in pairs that bind two opposing DNA target sites separated by a space. Hence they can recognize two DNA sequences of 9–18 bp (ZFNs) and 12–30 bp (TALENs), each separated by 5–7 bp. Meganucleases are characterized by a recognition site of 12–40 bp. Depending on the level of specificity of the SDN used and presence/absence of sequences homologous to the recognition site in the genome, off-target modifications at other sites in the genome can be expected to occur at some frequency.

## ***Mutagenesis***

The engineered nuclease is designed to generate a DSB at a specific site (gene) in the genome. Until effective methods can be developed to introduce the SDNs themselves into plants, the gene encoding the SDN needs to be stably inserted by transformation or transiently expressed. Furthermore, inducible expression may be required if constitutive expression has undesirable off-target effects. So typically, a transgene that produces the nuclease is transformed into an unlinked region of the genome. The transgene is expressed and the protein generates a DSB at the targeted locus, resulting in a heterozygous mutant ( $T_0$ ) plant. The DSB created by the nuclease is repaired by the host cell's non-homologous end joining (NHEJ) DNA repair mechanism which often results in base substitutions or small DNA insertions or deletions (indels) at the break site (mutations). The segregating (selfed)  $T_1$  plants are genotyped to identify individuals that are homozygous for the mutation and no

longer carry the transgene. However, whole-genome sequencing is required to confirm the absence of any transgene fragments in the resulting lines. Insertions or deletions introduced into nuclease target sites (break sites), located in a gene's open reading frame, commonly cause frame shift mutations with loss of gene function (a gene knockout).

### ***Gene Replacement or Editing***

The aim is to modify a gene rather than eliminate its function by using the homologous recombination (HR) pathway to repair the nuclease-induced DSB. This is achieved by supplying a DNA donor template that contains sequences homologous to those flanking the break site; albeit currently at a low frequency of about 1 % as NHEJ still operates. This strategy is more complex than mutagenesis in that the donor template has to be delivered to the cell at the same time the nuclease mediates cleavage. The HR pathway uses the homology in the donor template to repair the DSB, thereby incorporating the donor sequence into the chromosome. The donor sequence is designed to achieve the desired changes in the gene, either to the promoter or to the coding regions that alter the catalytic activity of an enzyme.

### ***Gene Insertion (SDN-3 Technique)***

The aim is to introduce a transgene or transgenes at a specific chromosomal location through HR. In practice, a recombinant DNA molecule is constructed in which the DNA fragment or the gene cassette of interest is sandwiched between stretches of DNA that are homologous with the DNA sequences flanking the DSB site. Sites of gene insertion are chosen that are conducive to high levels of gene expression. Furthermore, the insertion of multiple transgenes at the same chromosomal locus (stacking) makes it easy to move the transgenes into other germplasm by crossing, because the linked transgenes behave genetically as a single locus. EFSA (EFSA 2012b) concluded that the main difference between the SDN-3 technique and transgenesis is that the insertion of DNA is targeted to a predefined region of the genome. Therefore, the SDN-3 technique can minimise hazards associated with the disruption of genes and/or regulatory elements in the recipient genome. Whilst the SDN-3 technique can induce off-target changes in the genome of the recipient plant these would be fewer than those occurring with most mutagenesis techniques. Furthermore, where such changes occur they would be of the same types as those produced by conventional breeding techniques.

### ***Site-Directed Structural Changes***

Although perhaps less common, a breeder might want to induce specific structural changes to the genome. A nuclease or a combination of two nucleases is targeted to cleave neighbouring locations along a chromosome. Some of the breaks will be repaired by joining the respective outer breakpoints, thereby deleting the sequence separating the two cleavage sites. Alternatively, it is possible that the released DNA fragment will invert before DSB repair, thereby resulting in an inversion of sequence between the two cleavage sites. Similar methods can be used to stimulate chromosomal translocations if the two nucleases target sites on different chromosomes.

### ***Rice Example***

One of the first successful demonstrations of a nuclease-mediated modification of agronomic importance was resistance in rice to the phytopathogenic bacteria *Xanthomonas oryzae*, the causal agent of a devastating bacterial blight (Li et al. 2012). During infection the bacterium secretes effector proteins that target DNA sequences in the promoter region of the rice sucrose-efflux transporter gene (*OsSWEET14*). This binding transcriptionally activates *OsSWEET14*, which thereby contributes to pathogen survival and virulence. Since *OsSWEET14* plays an important role in the development of the plant, obtaining a knockout mutant to circumvent the effects of the pathogen was not feasible. The authors therefore transformed rice with a pair of TALENs designed to disrupt the promoter sequence bound by the effector, without disrupting the TATA box required for normal gene function. Several independent mutant lines were recovered, some of which showed resistance to *Xanthomonas oryzae*.  $T_1$  generation segregants were genotyped to identify plants that had maintained the site-directed mutation while segregating away the engineered nuclease transgene. Rice plants that were resistant to bacterial infection were shown to be either homozygous monoallelic or heterozygous biallelic mutants at the targeted site. Furthermore, resistant plants were morphologically identical to wild-type, indicating that the mutations had not resulted in adverse developmental phenotypes.

### ***CRISPR/Cas9 System for Plant Genome Editing***

The type II clustered regularly interspaced short palindromic repeat (CRISPR/Cas9) system from *Streptococcus pyogenes* is a technique for the introduction of site-specific double-stranded DNA breaks into genomes. It was discovered as part of the adaptive immune system of bacteria and archaea, protecting them against

invading nucleic acids such as viruses by cleaving the foreign DNA in a sequence-dependent manner. It has been developed over the period since 2005 and was first demonstrated in plants in 2013. Examples are now available in wheat, rice, maize, sorghum and tomato. Bortesi and Fischer (2015) have reviewed the results in plants and discussed possible applications in plant breeding. Some of their key points are as follows. The system consists of a single monomeric protein (Cas9) and a chimeric guide RNA (gRNA) which binds with the target DNA sequence. The chimeric RNA combines a CRISPR RNA (crRNA) for target specificity and a trans-activating CRISPR RNA (tracrRNA), which stabilizes the structure and activates Cas9 to cleave homologous double-stranded DNA sequences known as protospacers. Sequence specificity is conferred by a 20-nucleotide (nt) sequence in the gRNA of which the so-called seed sequence of approximately 12 nt is thought to be particularly important for the pairing between RNA and target DNA. The presence of a protospacer-adjacent motif (PAM) directly downstream from the target DNA is a prerequisite for DNA cleavage by Cas9. The PAM is either a 5'-NGG-3' sequence or less frequently a NAG sequence. Activity of the system is dependent on cell type and delivery method. Functional gRNA and Cas9 protein can be delivered directly into plant cells or cells can be transformed with gRNA and Cas9 transgenes on a separate chromosome to the targeted locus so that they can subsequently be removed by segregation. Off-target effects appear to be rare and the careful selection of specific gRNA sequences should minimize the risk of such unwanted genome modifications.

Genome editing can therefore accelerate plant breeding by allowing the introduction of precise and predictable modifications directly into an elite cultivar background. Multiple traits can be modified simultaneously because multiple gRNAs with different sequences can be used to achieve high-efficiency multiplex genome engineering at different loci simultaneously. NHEJ-mediated gene knock-outs are the simplest form of targeted modification. For example, Wang et al. (2014b) targeted the genes of the mildew-resistance locus (*MLO*) in hexaploid bread wheat and successfully knocked out all three *MLO* homoeoalleles, thus generating plants with broad-spectrum resistance to powdery mildew disease. Precise nucleotide exchanges using oligonucleotide donor sequences could be used to modify the regulatory sequences upstream of genes that determine agricultural performance with the prospect of improving crop yields. The insertion of large sequences by NHEJ or HR would allow the introduction of transgenes at defined loci that promote high-level transcription and do not interfere with the activity of endogenous genes. Site-specific nucleases also allow the addition of several genes in close vicinity to an existing transgenic locus. This makes it feasible to introduce multiple traits into crops with a low risk of segregation in subsequent hybridizations (i.e. they are inherited together). The potential to introduce transgenes at a specific and predetermined chromosomal position using site-specific nucleases should all but eliminate the risk of unpredictable events. Hence there is the prospect of plants altered by the excision of a few nucleotides using genome editing tools such as CRISPR/Cas9 not being classified as genetically modified organisms. One further application of CRISPR/Cas9 that is likely to expand in the future is the targeted

insertion of transgenes in the fields of metabolic engineering and molecular farming, where plants or plant cells are used as factories for the production of specific metabolites or proteins.

## Somaclonal Variation

Cell culture is a key technique in the use of biotechnology for crop improvement, especially for gene transformation. However, plants regenerated from cultured cells often show a high frequency of alternations in their phenotypes, particularly when there is a callus phase, a phenomenon known as somaclonal variation. Hence there is a risk that regenerated plants will carry background mutations for undesirable characteristics that need to be screened out of the crop improvement programme. In contrast, there have been examples of desirable traits that could be used in plant breeding. Examples of some useful variants generated as a result of somaclonal variation can be found in the review by Bairu et al. (2011), which also summarizes the possible causes of somaclonal variation and methods of detection. An understanding of the molecular nature and mechanisms of somaclonal variation is important for evaluating the background mutations in regenerated plants and for utilizing somaclonal variation efficiently in plant breeding.

Miyao et al. (2012) revealed the molecular spectrum of somaclonal variation in rice by analyzing the whole-genome sequences of three plants independently regenerated from cultured cells originating from a single seed stock. They found that many single nucleotide polymorphisms (SNPs) and insertions and deletions (indels) contributed to the somaclonal variation in addition to the transposition of the endogenous retrotransposon Tos17. The observed molecular spectrum was similar to that of the spontaneous mutations in *Arabidopsis thaliana*. However, the base change rate was estimated to be  $1.74 \times 10^{-6}$  base substitutions per site per regeneration, which is 248-fold greater than the spontaneous mutation rate of *A. thaliana*.

## Mutation Breeding Versus Transgenesis

Before we consider genetically modified (GM) crops in the next chapter, we will finish this one with an interesting comparison of the outcomes of mutation breeding and genetic modification in rice. Batista et al. (2008) used oligonucleotide microarrays to analyze gene expression in four types of rice plant against respective controls. The four types were: a stable (on self-pollination) mutant produced by gamma irradiation, the ‘unstable’ M<sub>1</sub> generation of a 100-Gy gamma irradiated plant, a stable (on self-pollination) transgenic plant obtained for production of an anticancer antibody, and the ‘unstable’ T<sub>1</sub> generation of a transgenic plant produced for abiotic stress improvement. The transcript profile of the stable lines was less

altered than that of the unstable ones and GM plants showed fewer genetic alterations than mutagenized ones. Although these results may be specific to the particular mutagenized and transgenic plants examined, they show that transgenic plants may have fewer changes than mutagenized ones. The authors concluded that the safety assessment of new cultivars should be done on a case-by-case basis and not simply restricted to foods obtained through genetic engineering. A more detailed assessment of such issues can be found in the EFSA scientific opinion addressing the safety assessment of plants developed using Zinc Finger Nuclease 3 and other Site-Directed Nucleases with similar function (EFSA 2012b).

# **Chapter 17**

## **Genetically Modified Crops**

### **Introduction**

Seed producing inbred lines (and hence their hybrids) and vegetatively propagated clones can be modified by genetic transformation to produce genetically modified (GM) crops, which are also referred to as biotech crops or transgenic crops. The process results in the insertion of one or more agriculturally important genes into the genome of the crop species and thus allows targeted improvements of successful and widely grown cultivars. The first GM crops were planted in 1996 and by 2014 occupied 181.5 million hectares, over 10 % of all cropland in the world (James 2014). Further statistics can be found in Chap. 2 including the fact that four crops (soybean, maize, cotton and canola) and two traits (herbicide tolerance and Bt insect resistance) dominate. Other GM crops being grown in 2014 were alfalfa, eggplant, papaya, poplar, squash, sugar beet, sweet pepper and tomato; making 12 commercial GM crops in total. James (2014) attempted to predict likely new GM crops and traits over the next 5–10 years. The crops were: apple, banana, camelina, cassava, citrus, chickpea, cowpea, groundnut, mustard, pigeon pea, potato, rice, safflower, sorghum, sugarcane and wheat. The traits included ones for better quality, new disease and pest resistances, tolerance of abiotic stresses such as drought and salinity, and improved agronomic performance including yield and nitrogen utilization.

In this chapter we are going to take a look at some GM crops and issues surrounding them. It will become clear that the prerequisites for successfully producing and marketing a GM crop are as follows: understanding the biochemistry of the trait of interest; understanding the control of gene expression for the genes that encode the key enzymes for delivering the required biochemistry; the ability to clone the key gene(s) and to do the necessary genetic engineering in bacterial plasmids to produce vectors containing transgene expression cassettes; the ability of the vector to deliver the expression cassette into plant cells; the ability to select or produce marker-free transformants; the ability to regenerate whole plants from

transformed cells and tissues; the ability to stack combinations of transgenes; demonstration that the resulting GM crop and products are safe to grow and eat or feed to animals; meeting all of the regulations governing GM crops; and last but possibly most difficult of all, convincing politicians, pressure groups and the general public that the GM crop is of real benefit to farmers and consumers alike. Perhaps it is not surprising that a 2011 Philips McDougal publication (James 2014) reported that the costs for discovery, development and authorization of a new plant biotechnology trait introduced between 2008 and 2012 was US\$136 million, and that the average time from initiation of a discovery project to commercial launch was 13.1 years.

The reader is referred to books on plant biotechnology for the details of gene cloning and genetic engineering of plasmids; for example: the second edition of *Plant Biotechnology the genetic manipulation of plants* by Slater et al. (2008). We are going to start with a few definitions and brief summaries of techniques and regulations governing the release and utilization of GM crops before considering some actual examples.

## Transformation Techniques

Most plant transformation systems require the ability to regenerate plants from isolated cells or tissues *in vitro*. Hence there is a need to find and combine a suitable plant tissue culture and regeneration regime with a compatible plant transformation protocol. Here, however, we are going to focus on the latter. The reader can find accounts of plant tissue culture in books such as the one by Slater et al. (2008). A prerequisite for transformation is the isolation and cloning of the desired gene. The reader is referred to the fourth edition of the book by Heldt and Piechulla (2011) for an account of how this is done, including production of a cDNA library from the mRNA present in the cellular source of the desired gene, cloning the cDNA into plasmids that are maintained in colonies of *Escherichia coli*, screening the colonies for the desired gene, and further genetic engineering of the colony with the desired gene.

*Agrobacterium*-mediated and direct gene-transfer methods have been developed for the stable transformation of plants. *Agrobacterium*-mediated transformation uses the natural ability of this bacterial pathogen of dicotyledonous (broad-leaved) plants to transfer genes into plant genomes. Direct gene-transfer methods rely on the delivery of large amounts of naked DNA while the plant cell is transiently permeabilized. Particle bombardment (biolistics) is the major direct gene-transfer method which has commonly been used for the transformation of monocotyledonous plants, particularly cereals. Other direct gene-transfer methods not considered here are electroporation of plant cells and protoplasts, treatment of protoplasts with polyethylene glycol (PEG) and use of silicon carbide fibres to penetrate the cell wall and plasma membrane. Descriptions of the techniques can be found in the book by Slater et al. (2008). Brief accounts follow of *Agrobacterium*-mediated

and biolistic-mediated transformation, with examples from potatoes and rice, respectively; and an interesting finding in sweet potato cultivars.

### ***Agrobacterium-Mediated Transformation***

*Agrobacterium tumefaciens* (*Rhizobium radiobacter*) is a soil-borne, Gram-negative, rod-shaped, motile bacterium that is the causative agent of crown-gall disease of crops such as apples and grapes. It is found around the roots of plants and normally survives on the nutrients they release. However, if the plant becomes wounded or damaged at the stem base, it can infect the plant at the wound site and cause disease symptoms. Crown-gall formation depends on the presence in *A. tumefaciens* of a Ti plasmid (tumour-inducing,  $\approx 200$  kb in size) whose T-DNA (transfer DNA,  $\approx 20$  kb in size) region is transferred from the bacterium into the plant cell and then into its nucleus, where it becomes integrated into the host's nuclear genome. The T-DNA has genes that encode proteins involved in hormone biosynthesis (auxin and cytokinin), the cause of cellular proliferation and gall formation (hence called oncogenes), and the production of opines (amino acid derivatives) and agropines (sugar derivatives) as a source of nutrients for the bacterium. Analysis showed that one or more copies of the T-DNA could be integrated into the host genome and that the insertions were bordered by small (24 bp) imperfect left- and right-border repeats which also border the T-DNA in the Ti plasmid. The borders in the plasmid are recognized by an enzyme complex and one of the enzymes (VirD2) produces single-stranded nicks in the plasmid DNA, and then becomes covalently attached to the 5' end of the displaced strand. Repair synthesis replaces the displaced strand. The T-DNA strand and associated proteins pass through the nuclear pore and the T-DNA strand is converted to double-stranded form. Integration into the host genome at naturally occurring chromosome (double-stranded) breaks occurs by non homologous end joining (NHEJ), which was explained in Chap. 3. As a consequence, there is no preference for integration of the donor genes into similar nucleotide sequences in the recipient genome; but the process may not be entirely at random. Different strains of *A. tumefaciens* exist and provide variations on this theme.

Procedures for *Agrobacterium*-mediated transformation were first published in 1984 by De Block et al. (1984) and Horsch et al. (1984). They were possible because the only features of the T-DNA required for integration into the host plant genome are the short border sequences; removal of the oncogene sequences enables plants to be regenerated from transformed plant tissue by manipulating the plant hormone composition of the medium; and the virulence genes in the plasmid responsible for the transfer of the T-DNA into the host plant cell function in *trans* (i.e. they could be on a separate plasmid). The techniques of molecular biology are used to convert the Ti plasmid into a vector for plant transformation. The details need not concern us here, but two points are worth making. Firstly, the initial isolation and cloning of the desired gene(s) and the genetic engineering of the

plasmid is commonly done in *E. coli* before transfer to *A. tumefaciens*. Hence the plasmid must be able to replicate in both these bacteria. Secondly, the genes that are going to be integrated into the genome of the host plant may come from prokaryotes or non-plant eukaryotes. Hence they made need to be given appropriate promoters and terminator sequences (to stop transcription) to ensure their expression in the host plant. Today binary vectors are commonly used in which the transfer apparatus (the virulence genes) and the T-DNA are located on separate plasmids. The former plasmid is maintained in a suitable strain of *A. tumefaciens* while the latter plasmid is normally maintained in *E. coli* and has the gene of interest (transgene) cloned into it. This plasmid can then be transferred into the *A. tumefaciens* strain for the plant transformation. Advances are continually being made in the design of plant transformation vectors and the regeneration of transformed plants. Furthermore, numerous factors can influence the level of donor gene expression in the plant, but the three most important ones are the characteristics of the genetic construct introduced, the number of copies inserted and the site of insertion. Hence it is important for breeders and biotechnologists to check the scientific literature for the methods currently being used for their own particular crop species.

### ***Agrobacterium*-Mediated Transformation of Potatoes**

Potato transformation using *Agrobacterium*-mediated systems was developed during the 1980s, first with *A. rhizogenes* (Ooms et al. 1986) and then more successfully with *A. tumefaciens* which today is the method of choice (Stiekema et al. 1988). The gene of interest would be incorporated into the bacterial plasmid along with a promoter and selectable marker, such as resistance to an antibiotic or herbicide; the bacterium would be co-cultured with freshly cut tuber discs or leaf or internode explants of the potato; and regeneration of shoots with the selectable marker would take place in plant tissue culture in the presence of the selectable agent to ensure selection of transformants. Regeneration could be a problem with some genotypes so the easily transformed cultivar Desiree was a frequent choice for research purposes (Davies 2002). Both Chang et al. (2002) and Morris et al. (2006) were able to simultaneously co-transform potatoes with two genes using a single selectable marker. The transgene constructs were multiplied separately in *A. tumefaciens* clones and cultures of the *Agrobacterium* were mixed and incubated with the potato internode explants. From 300 explants, Morris et al. (2006) generated 38 independent transformants of which 4 contained both transgenes. Clearly this increases the speed and efficiency of transformation as a breeding method and makes it a more attractive proposition. The choice of promoter is important for gene expression. The CaMV 35 S promoter has frequently been used for constitutive gene expression, but others have been developed for higher constitutive expression and for leaf or tuber expression (Douche and Grafius 2005). Selecting the best transgenic potato clones for commercialization required a lot of work; namely seeking the best constructs and promoters and eliminating any undesirable

transformants or somaclonal variants that arose from the transformation *per se* and *in vitro* regeneration, respectively. Davies (2002) recommended the production of several hundred independently transformed lines and full and effective field selection as normally done by plant breeders in their programmes. However, we shall shortly see that commercialization of GM potatoes encountered other difficulties. Hence the following discovery in sweet potato is of particular interest.

### ***Naturally Transformed Sweet Potato***

Kyndt et al. (2015) found that among 291 tested accessions of cultivated sweet potato, all contained one or more transfer DNA (T-DNA) sequences. Furthermore, they analyzed a cultivated sweet potato clone ("Huachano") in detail and showed that these sequences were expressed. One of the T-DNAs is apparently present in all cultivated sweet potato clones, but not in the crop's closely related wild relatives. The authors tentatively concluded that an *Agrobacterium* infection occurred in evolutionary times and that the T-DNA provided a trait or traits that were selected for during domestication. Sweet potato is therefore a widely and traditionally consumed food crop that is also a naturally transgenic crop, a discovery that could affect the current consumer distrust of the safety of transgenic food crops.

### ***Biostatic-Mediated Transformation***

Particle bombardment systems were first developed in 1987 (Klein et al. 1987) and used an explosive charge to propel double-stranded DNA-coated tungsten (later gold) particles at high speed into the target plant material, where the DNA was released within the cells and could integrate into the genome (Slater et al. 2008). Developments to the technology resulted in a number of systems of which a commercially produced, helium-driven, particle-bombardment apparatus (PDS-100/He) became widely used. The following brief description is taken from Slater et al. (2008). The plant tissue is placed into a vacuum chamber below a microcarrier stopping plate. The vector DNA-coated particles (the microcarriers) on a macrocarrier membrane are inserted above the stopping plate. Above them is a rupture disc. When the helium pressure in the apparatus above the rupture disc is increased, the disc bursts and this propels fragments of the macrocarrier and the projectiles down the chamber. The macrocarrier is stopped at the stopping plate while the microcarriers pass through and hit the plant tissue. The system needs to be optimized in terms of particle type and preparation, particle acceleration and choice of target material. Plant tissues used for bombardment are usually either primary explants that are bombarded and then induced to become embryonic, or proliferating embryonic cultures that are bombarded and then allowed to proliferate further and subsequently to regenerate. Bombardment has mainly been used for

transforming the nuclear genome but can also be used for introducing DNA into plastids for chloroplast transformation, where there is the potential for very high levels of gene expression and synthesis of active proteins. In many crop species chloroplasts display only maternal inheritance so that there is no danger of any gene transfer to related weedy species through pollen. If there are plastid sequences in common on the vector and in the plastid genome, integration is by (site-specific) homologous recombination. As with *Agrobacterium*-mediated transformation, it is important for breeders and biotechnologists to check the scientific literature for the biolistic methods currently being used for their own particular crop species.

### ***Biolistic-Mediated Transformation of Rice***

The example of biolistic-mediated transformation given by Slater et al. (2008) was use of the PDS-1000/He system for rice. Two plasmids coated onto gold particles were introduced into the plant cell together, one carrying the transgene of interest and the other both a selectable marker (antibiotic-resistance) and a reporter gene to assess gene expression and as an indicator of transformation. Slater et al. (2008) give details of how the plasmid DNA from *E. coli* is attached to gold microcarriers which are then applied to the macrocarrier membrane. They also give details of how embryogenic callus tissue is produced from dehulled, mature rice seeds. After bombardment, cells containing the antibiotic-resistance gene were selected on culture medium containing the antibiotic. In many cells the transgene also integrated into the plant genome. Plant regeneration was done *in vitro* to produce young plantlets with roots and shoots for transfer to soil and growth to maturity. Through self-pollination it was possible to segregate out the selectable marker, leaving transgenic plants with just the gene of interest.

Kohli et al. (1998) working with rice provided some insight into the integration process. They analyzed transgenic plants produced by particle bombardment (with two constructs) of 16 rice lines representing six popular cultivars. Their results indicated that multiple integration events had occurred as a cluster giving rise to a single transgenic locus in all of the lines analyzed. The multiple copies of transgenes were separated by plant genomic DNA. The authors proposed a possible two-phase mechanism for integration of transgenes. In the first phase penetration of cells with DNA-coated metal particles elicits a wound response which activates the enzymatic machinery for host DNA repair and foreign DNA degradation. This results in extrachromosomal rearrangements and ligations of the transforming plasmid molecules before integration. In the second phase initiation of integration of the scrambled and/or intact transgenic sequences occurs in competent cells at specific target sites. The original site of integration acts as a hot spot, facilitating subsequent integration of successive transgenic molecules at the same locus. The authors thought that naturally occurring double-stranded break points in the plant genome serve as integration sites through the mechanism of non homologous end joining (NHEJ).

## Clean-Gene Technology and Gene Stacking

Initially it was common practice to use antibiotic- or herbicide-resistant genes as markers for the selection of transformants. However, public concerns arose over the risk to human and animal health of using antibiotic-resistance marker genes. Today marker-free transformants can be produced, most easily by using transformation vectors without selection markers, followed by PCR analysis of regenerated plants for the presence of the transferred gene(s). We will consider such an example in the section on late blight resistant potatoes (Jo et al. 2014). However, if necessary, marker genes can be eliminated from the final plant by several techniques (EFSA 2012a, b), one of which we will encounter with HarvXtra™ alfalfa.

Our initial examples in this chapter will be fairly simple, single-gene transformations for traits such as herbicide tolerance, virus resistance and insect resistance. These single genes are now being combined (stacked) in new GM cultivars, as well as the genes required for new biosynthetic capacity such as the production of provitamin A in Golden Rice. There are four basic strategies for doing this: (1) hybrid stacking in which a homozygous plant containing one or more transgenes is crossed with another plant containing other transgenes and the offspring selected for the presence of all the transgenes, and multi-stack hybrid production by iterative hybridizations which can be labour-intensive and time consuming; (2) - co-transformation in which a plant is transformed with two or more independent transgenes in separate gene constructs which are delivered to the plant simultaneously; (3) linked genes or multigene cassette transformation in which a plant is transformed with a single construct that contains two or more linked transgenes, with 4–6 considered the current limit; and (4) re-transformation in which a plant containing a transgene is transformed with other transgenes. We will encounter all of these methods and James (2014) has summarized the stacked genes and traits being grown in 2014. The largest combination to date is eight genes coding for three traits in SmartStax™ maize, which was launched in the USA and Canada in 2010 by Monsanto and the Dow Chemical Company (James 2014). It was the result of crossing four different GM maize lines to combine two herbicide tolerance genes (tolerance to glyphosate and glufosinate with different modes of action) with six *Bt*-genes for resistance to lepidopteran and coleopteran insects (explained below). The hope is that it will prevent the establishment of resistances to herbicides and the *Bt*-proteins among weeds and insect pests, respectively (discussed below).

As mentioned earlier, the products of transformation require screening to select the best transformants for commercialization. Commercialization also requires the demonstration that GM crops and products are safe to grow and eat or feed to animals. Regulations governing GM crops, food and feed are also often designed to address any other public concerns and to allow farmer and consumer choice (labelling and traceability) over whether or not to grow GM crops and eat GM products.

## Regulations Governing GM Crops, Food and Feed

Regulations vary according to country and breeders and biotechnologists will need to be familiar with those governing the target environments of their proposed GM crops, as well as those governing their research and development in the laboratory, glasshouse and field. For example, in the USA and Canada, no separate or additional regulatory approval is necessary for commercializing hybrid stacks produced from already approved GM plants, whereas in the EU and Japan stacks are considered new events which require an assessment of additional risks that could arise from the combined genes. Here we will consider two contrasting sets of regulations, the very stringent ones for the European Union (EU) and the less stringent ones for the USA which focus on the product and not the technology used for production. An issue in risk assessment has been whether or not the ‘precautionary principle’ should apply to GM crops. Some argue that the complexity and novelty associated with GM technology mean that the only concern should be appropriate measures to anticipate and prevent harm while others think that the consequences of not growing a GM crop should also be considered.

### ***EU Legislation***

EU Directive 2001/18/EC (of the European Parliament and of the Council of Ministers of 12 March 2001) deals with the deliberate release into the environment of genetically modified organisms and replaced Council Directive 90/220/EEC. It sets out procedures for considering applications to release GMOs (Genetically Modified Organisms) into the environment for research or commercial purposes. The current definition of genetic modification in the European Union (EU) includes the techniques of inducing mutations and fusing protoplasts, but in Annex 1b of the EU Directive 2001/18/EC both techniques have been exempted.

Regulation (EC) No 1829/2003 (of the European Parliament and of the Council of Ministers of 22 September 2003) deals with genetically modified food and feed. It creates a specific harmonized procedure for the scientific assessment and authorization of GM food and feed products. It also requires labelling of all GM food and feed which contains or consists of GMOs, is produced from GMOs (like glucose syrup from GM maize), or contains ingredients produced from GMOs (e.g. tomato paste and lecithin from GM soya, which is used as an emulsifier in chocolate bars).

Regulation (EC) No 1830/2003 (of the European Parliament and of the Council of Ministers of 22 September 2003) concerns the traceability and labelling of genetically modified organisms and the traceability of food and feed products produced from genetically modified organisms and amends Directive 2001/18/EC. It sets out EU requirements for a document audit trail to account for and identify approved GM products throughout the marketing chain, with the aim of facilitating accurate labelling.

In England ([www.gov.uk](http://www.gov.uk)) the directive is implemented through the Environmental Protection Act 1990 (EPA) which gives the government general powers and responsibilities to control the deliberate release of GMOs in England. The Genetically Modified (Deliberate Release) Regulations 2002 supplements the EPA by setting out detailed rules for the implementation of Directive 2001/18/EC, including specific requirements for applications to release GMOs. Regulations 1829/2003 and 1830/2003 are implemented in England through the following regulations: Genetically Modified Food (England) Regulations 2004; Genetically Modified Animal Feed (England) Regulations 2004; and Genetically Modified Organisms (Traceability and Labelling) (England) Regulations 2004. Similar regulations have been implemented in Northern Ireland, Scotland and Wales. In this way, the UK as a member state of the EU implements the EU regulations. More details can be found in Slater et al. (2008). On 13 January 2015 members of the European Parliament passed a new law that will allow EU member states from spring 2015 to restrict or ban the cultivation of GM crops in their own territory, even if this is allowed at EU level, thus allowing greater flexibility within the EU.

## ***EFSA Guidance***

The European Food Safety Authority (EFSA) does scientific evaluations of potential GM crops and products for the European Commission and has produced useful guidance on the EU regulations. The guidance is rigorous and as comprehensive as possible. The European Commission (EC) is the executive body of the European Union responsible for proposing legislation, implementing decisions, upholding the EU treaties and managing the day-to-day business of the EU.

The 2010 document (EFSA 2010) produced by the GMO panel provides guidance for the environmental risk assessment (ERA) of genetically modified (GM) plants submitted within the framework of Regulation (EC) No. 1829/2003 on GM food and feed or under Directive 2001/18/EC on the deliberate release into the environment of genetically modified organisms (GMOs). In particular it provides guidance to risk assessors for assessing potential effects of releasing GM plants into the environment and the rationales for data requirements in order to complete a comprehensive ERA, and to draw conclusions for post-market environmental monitoring (PMEM). The document recommends that the ERA should be carried out on a case-by-case basis, following a step-by-step assessment approach with the six steps indicated in Directive 2001/18/EC: (1) problem formulation including hazard identification; (2) hazard characterization; (3) exposure characterization; (4) risk characterization; (5) risk management strategies; and (6) an overall risk evaluation. The document considers seven specific areas of concern to be addressed by applicants and risk assessors during the ERA. Briefly they are as follows.

1. Persistence and invasiveness of the GM plant, or transgenic wild relatives resulting from plant-to-plant gene transfer, within either agricultural or other production systems, or semi-natural and natural habitats.
2. Plant-to-micro-organism gene transfer (integration into genome of micro-organism) as a result of recombinant DNA released from GM plants into the environment; for example into soil or inside the gut of animals feeding on plant material.
3. Interaction of the GM plant with target organisms (generally pests or pathogens) on which specifically designed characteristics of a GM plant are intended to act.
4. Interaction of the GM plant with non target organisms, including criteria for selection of appropriate species and relevant functional groups for risk assessment of the potential environmental impact on population levels of herbivores, natural enemies, symbionts (where applicable), parasites, and pathogens.
5. Impact of the specific cultivation, management, harvesting and processing techniques associated with the GM plant, including consideration of the production systems (agriculture, forest tree or others) and the receiving environment(s).
6. Effects on biogeochemical processes in both the production site and the wider environment. Biogeochemical processes include the uptake of carbon dioxide from the atmosphere by plants, degradation of plant material, formation of soil organic matter, evaporation of water from fields, and transformation of nitrogenous compounds.
7. Effects on human and animal health; for example, exposure to pollen or dust from processed plants.

The 2011 document (EFSA 2011) produced by the GMO panel provides updated guidance for the risk assessment of food and feed containing, consisting or produced from genetically modified (GM) plants, submitted within the framework of Regulation (EC) No 1829/2003 on GM food and feed. The proposed risk assessment strategy for GM plants and derived food and feed seeks to deploy appropriate approaches to compare GM plants and derived food and feed with their respective comparators. The underlying assumption of this comparative approach is that traditionally cultivated crops have gained a history of safe use for consumers and/or domesticated animals. More generally, the safety of novel foods and food products has been assessed in terms of whether the novel food product is substantially equivalent to an existing analogous food product, but the application of substantial equivalence to GM foods has been criticized (Slater et al. 2008). The substantially equivalent approach was strengthened in the new regulations (1829/2003) by requiring more evidence.

The EFSA document provides guidance on how to perform the comparative analysis of the relevant characteristics of the GM plant. The document addresses the details of the different components of the risk assessment which are as follows.

1. Molecular characterisation, which provides information on the structure and expression of the insert(s) and on the stability of the intended trait(s).

2. Toxicological assessment, which addresses the impact of biologically relevant change(s) in the GM plant and/or derived food and feed resulting from the genetic modification.
3. Assessment of potential allergenicity, of the novel protein(s) as well as of the whole food derived from the GM plant.
4. Nutritional assessment to evaluate whether or not food and feed derived from a GM plant is nutritionally disadvantageous to humans and/or animals.

In addition every section of the document addresses specifically the requirements for GM plants containing a combination of transformation events, providing guidance on how to establish that the combination is stable and that no interactions occur between the events that may raise safety concerns.

## ***USA Legislation***

In the USA, regulations are implemented by three federal agencies: the US Department of Agriculture (USDA), the US Environmental Protection Agency (EPA) and the US Food and Drug Administration (FDA). The nature of the GM modification and its intended use determine which agency or agencies are involved in its regulation. Some US states require additional regulation of GM crops.

The Animal and Plant Health Inspection Service (APHIS) of USDA, through the Federal Plant Pest Act, determines whether a transgenic crop is likely to be an agricultural or environmental pest. Field tests of GM crops can often be conducted if APHIS is simply notified and certain criteria regarding safety of the GM crop are met, provided the applicant meets certain standards to ensure biological confinement. Otherwise a more detailed and involved process is required to obtain a permit. On the other hand, GM crops for commercialization can be granted non-regulated status by APHIS after presentation of extensive data on the nature of the genetic modification and its effects on the crop and the wider ecosystem. The GM cultivar and any offspring can then be released in the USA without APHIS review, but APHIS can prevent the continued use of a GM crop if it is considered a risk of becoming a pest.

The EPA, through the Federal Insecticide, Fungicide and Rodenticide Act, is responsible for ensuring the safety of pesticides including those produced biologically, such as plants genetically engineered to be pest-resistant. The EPA regulates the testing and commercialization of pest-resistant GM crops and sets tolerances for herbicide residues on herbicide-resistant crops.

The FDA, through the Federal Food, Drug and Cosmetic Act, regulates foods and animal feed derived from GM crops. Foods from GM crops have to meet the same safety standards as all other foods, but do not need FDA pre-market approval if their GM products are substantially equivalent to those in other foods.

## Cisgenic, Intragenic and Transgenic Crops

A distinction is made between transgenic, intragenic and cisgenic crops. The distinction lies in the source of the genes as cisgenesis and intragenesis make use of the same transformation techniques as transgenesis. Transgenic crops contain genes (transgenes) derived from organisms (e.g. bacteria and fungi) or other plant species that are not cross-compatible with the cultivated crop, whereas intragenic and cisgenic crops do not contain such genes. The following definitions have been used by the EFSA Panel on Genetically Modified Organisms (EFSA 2012a and online [www.efsa.europa.eu/efsajournal](http://www.efsa.europa.eu/efsajournal)). The additional comments have also been taken from this source.

“Cisgenesis is the genetic modification of a recipient organism with a gene from a crossable, sexually compatible, organism (same species or closely related species). This gene includes its introns and is flanked by its native promoter and terminator in the normal sense orientation. Cisgenic plants can harbour one or more cisgenes, but they do not contain any parts of transgenes or inserted foreign sequences. To produce cisgenic plants any suitable technique used for the production of transgenic organisms may be used. Genes must be isolated, cloned and transferred back into a recipient where stably integrated and expressed.” This definition implies the use of marker-free transformation. Sometimes the term cisgenesis is also used to describe an *Agrobacterium*-mediated transfer of a gene from a crossable, sexually compatible plant where T-DNA borders may remain in the resulting organism after transformation (cisgenesis with T-DNA borders). The T-strand transferred from *Agrobacterium* to the plant normally contains three nucleotides from the right border (RB) repeat and up to 22 nucleotides from the left border (LB) repeat, but their deletion can occur during integration. Database similarity searches indicate that sequences similar to the 22-bp LB repeat can be found in plants. Hence any hazards related to these sequences would not differ from those in conventional plant breeding. It is also possible for part of the vector backbone to remain attached to the border and to be co-inserted into the plant genome. This needs to be checked in any transformant as by definition cisgenic (and intragenic) plants must not contain vector backbone sequences of bacterial origin.

“Intragenesis is a genetic modification of a recipient organism that leads to a combination of different gene fragments from donor organism(s) of the same or a sexually compatible species as the recipient. These may be arranged in a sense or antisense orientation compared to their orientation in the donor organism. Intragenesis involves the insertion of a reorganised, full or partial coding region of a gene frequently combined with another promoter and/or terminator from a gene of the same species or a crossable species.” Again, the definition implies the use of marker-free transformation. Insertion has been done, in potatoes, using plant-derived transfer DNAs (P-DNAs) which resemble *Agrobacterium* T-DNA borders that support DNA transfer from *Agrobacterium* to plant cells (Rommens et al. 2007). Intragenesis includes the use of gene silencing approaches such RNA

interference which introduces inverted DNA repeats, as explained in the next section.

Some researchers have argued that cisgenesis with only one or a few cisgenes should be exempted from the definition of genetic modification (Jacobsen 2007). This is because the transformation process creates a GM plant whose genetic content is within the gene pool used by classical plant breeding, as also happens with induced mutations and protoplast fusion, both of which are exempt. The EFSA Panel on Genetically Modified Organisms did a safety assessment of plants developed through cisgenesis and intragenesis and published their opinion in 2012 (EFSA 2012a and online [www.efsa.europa.eu/efsajournal](http://www.efsa.europa.eu/efsajournal)). The Panel concluded that similar hazards can be associated with cisgenic and conventionally bred plants, while novel hazards can be associated with intragenic and transgenic plants. For example, the insertion of DNA fragments into plant genomes will create new open reading frames (ORFs) which are new nucleotide sequences uninterrupted by the presence of a stop codon. Potentially new sequence combinations, and hence novel proteins, could arise in transgenic and intragenic plants which would not normally arise in conventionally bred and cisgenic plants. The Panel concluded that their previous *Guidance for risk assessment of food and feed from genetically modified plants* (EFSA 2011) and *Guidance on the environmental risk assessment of genetically modified plants* (EFSA 2010) are applicable for the evaluation of food and feed products derived from cisgenic and intragenic plants and for performing an environmental risk assessment and do not need to be developed further. The Panel envisaged that on a case-by-case basis lesser amounts of (transformation) event-specific data may be needed for the risk assessment. One further point made by the Panel and worth repeating here is as follows. Cisgenesis and intragenesis do not introduce additional genes associated with linkage drag in introgression breeding from wild relatives, and thus avoid introducing any unwanted traits and hazards associated with these other genes. In introgression breeding, any potential cultivars producing such undesirable phenotypes are removed by breeders during testing and selection.

## Gene Silencing

RNA interference (RNAi) and post-transcriptional gene silencing (PTGS) were first observed in plants when it was found that interactions between multiple foreign transgenes or between transgenes and endogenous genes led to the degradation of specific mRNAs. Details can be found in Slater et al. (2008). Briefly, the degradation was sequence-specific and resulted in the silencing of transgenes and endogenous genes. It was subsequently shown that RNA silencing and related gene-silencing phenomena depended on the production of small (21–26 nucleotides) double-stranded RNA molecules as part of an RNA degradation pathway. These small interfering RNAs (siRNAs) are produced from long, double-stranded RNA molecules which can arise from several different mechanisms. For example, in

transgene-induced gene silencing the product of transcription from antisense transgenes can bind to the complementary mRNA from the sense orientated endogenous gene. Whatever the mechanism, the long, double-stranded RNA is cleaved by endonucleases to yield multiple siRNA duplexes of two sizes and ultimately results in cleavage of target RNAs. Whilst unintended gene silencing presents problems for genetic modification that need to be overcome in vector design, vectors can also be deliberately designed for gene silencing as we saw for potatoes. This is done by producing vectors that directly or indirectly lead to the production of double-stranded RNA. The simplest RNAi method is to produce antisense RNA by placing the sequence to be expressed in the reverse orientation with respect to the promoter. The promoter directs the production of antisense RNA which is complementary to the endogenous mRNA and hence they form double-stranded RNA. However, a more effective way of directing RNAi is by producing a vector in which the transgene sequence has been cloned as an inverted repeat with the repeat units separated by a spacer (usually an intron) sequence, and expression is driven by a strong promoter. In other words the order is promoter, then sense sequence, then spacer, and finally antisense sequence. Transcription of the inverted repeat allows the two complementary regions to form a double-stranded hairpin RNA which is extremely efficient in RNAi-induced destruction of any homologous mRNA.

It is worth pointing out that microRNAs (miRNAs) are short (21–24 nucleotides) RNAs which are involved in the post-transcriptional gene silencing of endogenous RNAs. They are encoded by the plant genome and exhibit a high degree of complementarity to the endogenous mRNAs which they silence by cleavage. They are formed as duplexes in the nucleus and then exported to the cytoplasm where they are processed into a single strand that interacts with the complementary target mRNA. When discovered they were a novel class of RNA involved in regulating gene expression. All RNA silencing pathways are capable of directing the methylation of DNA and thereby silencing gene expression at the transcriptional level.

## Herbicide Tolerance

It is now time to look at some examples of current and future GM crops, starting with herbicide tolerance. This was one of the first traits to be genetically engineered because a lot was known about the modes of action of herbicides, the targeted biochemical pathways in plants, and possible single-gene mechanisms of tolerance (see Chap. 5). Furthermore, there was the commercial incentive for agrochemical companies to produce crops tolerant to the herbicides which they manufactured. Tolerance of a crop to a particular herbicide allowed it to be treated at the optimal times for the reduction of weeds. Strategies and details for engineering herbicide tolerance can be found in Slater et al. (2008). Here we will look briefly at glyphosate tolerance as an example. Glyphosate is a broad-spectrum herbicide

that is marketed as Roundup by Monsanto. It is a simple derivative of glycine that acts as a very effective competitive inhibitor of the enzyme 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS) because it resembles phosphoenolpyruvate (PEP). EPSPS catalyzes the reaction between phosphoenolpyruvate and shikimate-3-phosphate to form 5-enolpyruvylshikimate-3-phosphate, and this is a key step in the synthesis of the aromatic amino acids tryptophan, tyrosine and phenylalanine. Hence glyphosate acts as an herbicide by inhibiting the biosynthesis of aromatic amino acids and other products of the shikimate pathway.

Tolerant cotton, oilseed rape, soybean and sugar beet crops (Roundup Ready) have been engineered using a construct comprising a resistant EPSPS gene from *A. tumefaciens* fused to a chloroplast transit peptide (from *Arabidopsis* or petunia) and an enhanced CaMV 35 S promoter. The resistant gene encodes an EPSPS enzyme that retains its affinity for PEP but has a reduced affinity for glyphosate. The transit peptide ensures delivery of the EPSPS to its site of action in the chloroplast. The GM plant's own enzyme will still be functional in the absence of the herbicide and expression of both genes has been reported to increase lignin content and cause undesirable stem splitting in Roundup Ready soybeans during hot weather (Slater et al. 2008). Tolerant maize crops (Roundup Ready) contain a construct with a resistant EPSPS gene from maize fused to a rice promoter and maize chloroplast transit peptide sequence. The resistant gene was isolated by selection in tissue culture following mutagenesis. The construct was optimized for a monocotyledonous crop, and this emphasizes the point that a lot of work is required to translate ideas into efficient and effective constructs. The search is likely to continue for better herbicide tolerance with no undesirable side effects. Slater et al. (2008) describe alternative strategies for glyphosate tolerance based on a specific detoxification mechanism which degrades the herbicide, as well as ways of engineering tolerance to other herbicides such as glufosinate (phosphinothrinicin). Glufosinate is a competitive inhibitor of glutamine synthase and hence causes ammonium toxicity and inhibition of photosynthesis. Tolerance has been engineered through detoxification by phosphinothrinicin acetyltransferase.

### ***Herbicide Tolerant Crops in the USA 1996–2012***

GM crops have been grown in the USA since 1996 with tolerance to herbicides one of the main modifications. Herbicide tolerant (HT) soybeans were first grown commercially in 1996, followed by maize and cotton in 1997, canola in 1999 and sugar beet in 2007; in 2013 they accounted for 88 % of the total US plantings of these five arable crops (Brookes 2014). Brookes (2014) has analyzed the rapid and widespread adoption of these crops and come to the following conclusions. The economic and environmental (EIQ indicator) benefits to farmers amounted to \$21.7 billion of additional farm income and a 225 million kg reduction in herbicide active ingredient use over the period 1996 to 2012, relative to conventional production methods. Farmers also derived the economic and environmental benefits associated

with switching from a plough-based to a no tillage or conservation tillage production system. A broad range of mostly selective herbicides were replaced by one or two broad-spectrum herbicides (mostly glyphosate, but also glufosinate on canola), used in conjunction with one or two other complementary herbicides (in terms of active ingredient) to counter weed populations developing resistance to herbicides. This was a problem faced by all farmers, not just those growing GM crops. As a consequence, since the mid-2000s the average amount of herbicide applied and the associated environmental load have increased on both GM and conventional crops. Despite this increase in herbicide, tolerant crops continue to deliver significant economic and environmental gains to US farmers. The reader is referred to the paper by Brookes (2014) for separate analyses of the five crops.

## Virus Resistance in Papaya

Beginning in 1992 a serious outbreak of papaya ringspot virus (PRSV) reduced Hawaii's papaya (*Carica papaya*) production by over 40 % within 5 years, as various control measures failed. PRSV is an aphid-transmitted non-persistent (hence quickly transmitted) Potyvirus whose genome comprises a single-stranded positive sense RNA which acts as mRNA directly. It can also be transmitted by mechanical inoculation. Researchers at Cornell University and in Hawaii produced a completely resistant transgenic papaya line 55-1 by transforming commercial cultivar Sunset. The details can be found in the paper by Fitch et al. (1992). Briefly, the research involved isolating and sequencing the coat protein (CP) gene of PRSV from Hawaii, microprojectile-mediated transformation of 2,4-D-treated immature zygotic embryos of Sunset with an appropriate plasmid construction, selecting and regenerating plants transformed with the CP gene of PRSV, screening transformed plants for resistance to PRSV, field trials, deregulation and commercialization. The engineered trait could be stably transmitted genetically by seed. GM cultivars SunUp and Rainbow were rapidly developed and became available to farmers in 1998. Papaya production recovered in less than 4 years and transgenic papaya now accounts for about 75 % of the acreage on Hawaii. In 2011 Japan approved the import of transgenic papaya from the US (James 2014). SunUp is homozygous for the coat protein gene and is resistant to a range of isolates including ones from outside Hawaii. This is believed to be due to a gene-dosage effect leading to enhanced post-transcriptional gene silencing. Rainbow is hemizygous for the transgene as it is an  $F_1$  hybrid between SunUp and a non-transgenic line, Kapoho. As a result it is more sensitive to non-Hawaii isolates, but in Hawaii it is the cultivar preferred by the market and the growers. At present there is still a requirement for non-transgenic plants for some overseas markets as the transgenic plants are tagged with the  $\beta$ -glucuronidase (GUS) protein gene for ease of detection (Slater et al. 2008). The possible mechanisms of CP mediated resistance are presented and discussed by Slater et al. (2008).

## Insect Resistance

### Cry Endotoxin Genes from *Bacillus thuringiensis*

The following brief account of the use of *Cry* endotoxin genes from *Bacillus thuringiensis* (*Bt*) is taken from Slater et al. (2008) where more detailed information can be found. *B. thuringiensis* was discovered by Shigetane Ishiwatari in 1901 in diseased silkworms. The toxic effect of the bacteria on insect larvae was subsequently shown to be due to insecticidal crystal (*Cry*) proteins ( $\delta$ -endotoxins) which form inclusion bodies during sporulation. After ingestion by an insect larva, the crystals dissolve in its midgut and bind to receptors in the epithelium cells, killing them as a consequence by osmotic lysis. The toxicity of the crystals to other animals including humans is extremely low. The various strains of *B. thuringiensis* produce a wide range of different *Cry* proteins which share a common active core comprising three domains. They are encoded by *cry* genes carried on plasmids. Individual *Cry* proteins are active against the larvae of particular classes of insect. Preparations of *B. thuringiensis* spores or isolated crystals have been used as 'organic' pesticides for over half a century and this use aided the rapid development of the *Bt*-strategy based on genetic modification. However, the prokaryotic gene sequences of the *cry* genes had to be modified to obtain high levels of stable expression in plants. The specificity of *Cry* proteins allowed the targeting of specific pests by particular transgenes.

Cotton (*Gossypium* spp.), maize and soybean have been genetically engineered to produce *Cry* proteins and since 1996 *Bt*-crops have been planted on a cumulative total of more than half a billion hectares worldwide. Their deployment has reduced the use of broadly toxic insecticides and increased farmers' profits. However, the rapid evolution of resistance to *Bt*-toxins by some pests has reduced the benefits of this approach. To delay the evolution of resistance, farmers plant refuges of insect host plants that do not make *Bt*-toxins. These allow survival of insects that are susceptible to the toxins. When refuges near *Bt*-crops produce many susceptible insects, they reduce the chances that two resistant insects will mate and produce resistant offspring. The required refuge area of 20 % of GM crop has been reduced to 5 % in the US Corn Belt for SmartStax<sup>TM</sup> maize, with its six *Bt*-genes for resistance to lepidopteran and coleopteran insects. However, Gassmann et al. (2014) have advised against this reduction if the *Bt*-crop does not achieve a high dose of toxin against a target pest (i.e. kills 99.99 % of susceptible insects). This is the situation with currently commercialized *Bt*-maize that relies on *Cry3Bb1*, *mCry3A* and *Cry34/35Ab1* proteins to target the western corn rootworm. *Cry3Bb1* maize was registered in 2003 by the United States Environmental Protection Agency (US EPA) for management of western corn rootworm larvae (*Diabrotica virgifera virgifera*). In 2009, Iowa farmers observed severe injury to *Cry3Bb1* maize which subsequent laboratory assays revealed was associated with *Cry3Bb1* resistance. The injury to *Cry3Bb1* maize persisted through 2011 and expanded to include *mCry3A* maize. Analysis revealed that cross-resistance was

present between the closely related Cry3Bb1 and mCry3A proteins. The results demonstrated that insects can evolve resistance to *Bt*-crops that are not high dose in three generations (i.e. over 3 years of *Bt*-maize cultivation) and raised concerns about the adequacy of current resistance management strategies. Beginning in 2009, the US EPA approved *Bt*-maize with a pyramid of two *Bt*-toxins targeting the western corn rootworm, including Cry3Bb1 with Cry34/35Ab1 and mCry3A with Cry34/35Ab1. However, the presence of resistance to one toxin in a pyramid diminishes the effectiveness of a pyramid to delay resistance, and coupled with reduced refuge size, may hasten the evolution of resistance. The authors concluded that the 5 % refuge associated with current *Bt*-pyramids targeting western corn rootworm may do little to delay resistance, and that larger refuges should be considered as a tactic to delay resistance along with crop rotation to interrupt continuous maize fields.

### ***Insect Resistant Cotton in China***

In northern China there are ten million farmers and insecticidal transgenic *Bt*-cotton is grown as part of a vast patchwork of small farms. Although *Bt*-cotton accounts for 98 % of all cotton, cotton represents only 10 % of the area planted with crops eaten by the cotton bollworm (*Helicoverpa armigera*). Hence the deliberate planting of non *Bt*-cotton refuges was considered unnecessary. Jin et al. (2014) tested this natural refuge strategy by comparing predictions from modelling with data from a 4-year field study of cotton bollworm resistance to transgenic cotton in six provinces of northern China. The transgenic cotton produced a high dose of *Bt*-toxin Cry1Ac. Bioassay data revealed that the percentage of resistant insects increased from 0.9 % in 2010 to 5.5 % in 2013. The percentage of these resistant insects with non recessive inheritance of resistance (so that resistant × susceptible insects can produce some resistant offspring) increased from 37 % in 2010 to 84 % in 2013. Modelling predicted that the percentage of resistant insects would have exceeded 98 % in 2013 without the natural refuges, but would have increased to only 1.1 % if they had been as effective as non *Bt*-cotton refuges. The authors therefore concluded that natural refuges delayed resistance, but were not as effective as an equivalent area of non-*Bt* cotton refuges. They recommended switching to *Bt*-cotton producing two or more toxins and integrating other control tactics, such as biological control by predators and parasites, to slow further increases in resistance.

### ***Insect Resistant Cotton in India***

*Bt*-cotton is also grown in India with 95 % adoption. Krishna et al. (2015) showed that varietal diversity was reduced in the early adoption phase when only a small

number of *Bt*-varieties were approved but recovered when more transgenic varieties became available. Hence transgenic crops need not cause agrobiodiversity erosion.

### **Drought Stress Response of Insect Resistant Maize in Europe**

A major pest of maize worldwide is the European corn borer (*Ostrinia nubilalis*) whose larvae damage maize crops by tunnelling into the central pith of the stalks and ears. Genetically engineered resistance was one of the early targets of the agricultural biotechnology industry and one of the successful transformation events with the *cry1Ab* gene was MON810 from Monsanto, which they commercialized under the tradename YieldGard. Interestingly, it is currently the only cultivated GM event in the EU. It contains a stable, genome-integrated plant expression cassette (construct) comprised of the cauliflower mosaic virus 35 S constitutive promoter and HSP70 maize intron sequence, driving the expression of a truncated *cry1Ab* gene with modified codon usage to optimize its expression in maize. Gullì et al. (2015) compared maize cultivar DKC6575 (event MON810) with its near-isogenic non GM cultivar Tietar in a field and growth chamber for their response to drought. The *cry1Ab* transcript was not influenced by drought stress, but was expressed at a constant level in all of the experiments. Furthermore, the main photosynthetic parameters were affected to a similar extent by drought stress in the GM and non-GM cultivars. However, gene expression profiling under drought stress revealed that more stress-responsive genes were up-regulated in Tietar than DKC6575, indicative of a more efficient drought response in Tietar. Hence both predicted and unexpected changes can occur on genetic transformation.

It should be mentioned here that it was a different transformation event Bt176 that led to the Monarch butterfly affair. The Bt176 event produced Cry1Ab toxin in pollen at much higher levels than event MON810, with consequences for the potential toxicity of the pollen to non-pest lepidopterans. A controversial study from Cornell University published in 1999 had found that the larvae of Monarch butterflies fed on milkweed leaves (their host) covered in pollen from Bt176 maize did not grow as well as those on control leaves. Subsequently, six research groups did more work and published papers in 2001 in the *Proceedings of the National Academy of Sciences USA* (Scriber 2001). Readers may like to consult these papers and draw their own conclusions, but Bt176 is no longer available in US maize cultivars, so the overall conclusion is that the risk to Monarch butterfly populations in the USA is low from current GM cultivars. Slater et al. (2008) also examined plant insecticidal genes that have been used experimentally to engineer pest resistance, such as the cowpea trypsin inhibitor. In the UK it was the controversial work of Pusztai in 1998 (Ewen and Pusztai 1999) that precipitated a public backlash against GM food. He found that feeding transgenic potatoes carrying the snowdrop lectin GNA (*Galanthus nivalis agglutinin*) gene to rats resulted in unexpected (the natural lectin was not harmful to rats) and undesirable effects such as cell

proliferation of the gastric mucosa (i.e. stomach damage), but a number of critiques found the study inadequate (Slater et al. 2008).

## Drought Tolerant Maize

Having considered drought stress response of insect resistant maize in Europe, it is timely to consider drought tolerance itself. The first GM drought tolerant maize was planted in 2013 by US farmers, and as mentioned in Chap. 2, Monsanto have donated Event MON 87460 to Water Efficient Maize for Africa (WEMA). WEMA is a public-private partnership designed to deliver the first stacked biotech insect resistant and drought tolerant maize hybrids to selected African countries starting in South Africa in 2017 (James 2014). Africa grows 90 % of its maize under rainfed conditions and up to 25 % of the area suffers from frequent droughts. Event MON 87460 was the outcome of research done by Monsanto scientists in which they demonstrated that constitutive expression of two members of a family of bacterial RNA chaperones (cold shock proteins), *E. coli* CspA and *Bacillus subtilis* CspB, conferred abiotic stress tolerance in transgenic *Arabidopsis*, rice, and maize (Castiglioni et al. 2008). Cold shock proteins were discovered and named through their rapid accumulation in cold shocked bacterial cells. Their chaperone function of binding and unfolding tangled mRNA molecules so that they can function normally, is thought to be important for stimulating growth following stress acclimation and during periods of high metabolic activity, but the details are still not fully understood. In maize the positive yield impact of the transgenes was predominantly on kernel numbers, not weight, as expected from the target period for the water-deficit of 10–14 days prior to flowering. Matt DiLeo (*Biology Fortified*, 25 August 2012) describes the subsequent work done by Monsanto to bring their research to fruition. Maize was transformed with *cspB* and assessed in a series of field experiments of increasing complexity and realism. From 22 independent *cspB* transgenic events that were crossed into commercially acceptable maize, the most promising line, CspB-Zm event 1 (details can be found in plant pest risk assessment), was assessed in several years of field trials. The transgenic locus was crossed into three different hybrid genotypes, which were tested under three controlled watering conditions (well-watered, drought immediately preceding flowering, drought during grain fill) at five replicated locations. Non-transgenic controls suffered 50 and 30–40 % yield losses under the two drought stresses, respectively, while *cspB* lines produced 11–21 % relative yield gains. The yield stability of ‘CspB-Zm event 1’ was also analyzed over 4 years of testing in three hybrid test-crosses, where *cspB* provided an average yield benefit of 10.5 % (range of 6.7–13.4). *CspB* lines were also tested under dryland (non-irrigated) conditions at normal commercial planting densities. In the sites that experienced seasonal drought, the maximum yield gain of the transgenic lines was 15 %.

Drought tolerance is a complex physiological trait, so not surprisingly more than one way may be found of achieving tolerance, including through conventional

breeding. Hence a breeder needs to keep an open mind and explore various alternative strategies. In an article in *Nature* in 2014, Gilbert (2014) concluded that genetic engineering lagged behind conventional breeding in efforts to create drought-resistant maize. The Drought Tolerant Maize for Africa project, which was launched in 2006 with US\$33 million, had developed 153 new cultivars to improve yields in 13 countries. In field trials, these new cultivars matched or exceeded the yields from current cultivars under good rainfall conditions, and yielded up to 30 % more under drought conditions. The project's success was due in large part to its access to the seed bank at the International Maize and Wheat Improvement Centre (CIMMYT) in Mexico City. Breeders from CIMMYT and the International Institute of Tropical Agriculture (IITA) in Ibadan, Nigeria, searched the collection for maize cultivars that thrive in water-scarce regions. The researchers cross-bred these cultivars and then selected and hybridized the most drought-tolerant of their offspring. Several cycles of this process led to seed that was better adapted to water-scarce conditions. In a final step, project scientists cross-bred these plants with cultivars that have been successful in Africa. Effective selection was aided by the finding that resistance to drought was associated with shorter intervals between pollen release and silk emergence. When water is scarce, the silks emerge late and if the delay is long enough, they emerge after the plants have released their pollen and are not fertilized. There is of course no reason why conventionally bred cultivars cannot be further improved by genetic transformation; the two approaches are not mutually exclusive.

## Transgenic Drought Resistance and Tolerance

The physiological complexity of drought resistance and tolerance is brought home in a meta-analysis by Blum (2014) of 520 reports published during the last 20 years on transgenic and mutant plants generated to impart drought resistance. At least 487 tested transgenic plants involving at least 100 genes claimed to contribute to drought resistance and the number of experimental transgenic plants, both model and crop, for drought resistance has been impressive. However, there appears to have been very limited impact on global dryland agriculture and as we have seen, few drought resistant GM cultivars are currently being marketed. Blum argues that insufficient phenotyping of experimental transgenic plants for drought resistance often does not allow true conclusions to be made about the real function of the discovered genes towards drought resistance. He proposes an outline of a minimal set of tests that might help us resolve the real function of discovered genes, thus bringing the research results down to earth!

## Reluctance to Accept Genetically Modified Potatoes

### *Monsanto's Programme in the USA*

Monsanto were the first to commercialize GM potatoes, in North America from 1995. An account can be found in the review by Davies (2002). Briefly, processing cultivars Russet Burbank, Atlantic and Superior were transformed with a gene that encodes the Cry3A protein from *Bacillus thuringiensis*, under the control of the constitutive 35 S CaMV promoter. This confers resistance to the potentially devastating Colorado potato beetle (CPB, *Leptinotarsa decemlineata*) whose larvae feed on potato foliage. The cultivars were marketed under the tradename NewLeaf™ (e.g. Newleaf™ 6 Russet Burbank) in the USA from May 1995 when the USDA, EPA and FDA approved and de-regulated the product based on substantial equivalence to other Russet Burbank potatoes. Subsequently NewLeaf™ Plus was approved in August 1997. It is a high yielding Russet Burbank with combined resistance to CPB (*cry3A* gene) and *Potato leafroll virus* (PLRV). The latter was achieved using the constitutive Figwort Mosaic Virus (FMV) promoter within a construct designed to prevent virus replication. Finally NewLeaf™ Y cultivars Russet Burbank and Shepody were approved in May 1999. They have combined resistance to CPB and *Potato virus Y* (PVY); the latter achieved by using the FMV promoter to express the PVY coat protein gene. The steps from initial clone production (transplanting to soil) to commercialization of NewLeaf™ Plus and NewLeaf™ Y involved extensive evaluation and reduced the number of transformants from 3000 of each in 1991 to 6 in 1998. Despite all of this evaluation, leading processing and fast-food outlet companies in North America were reluctant to purchase GM potatoes, because of consumer concerns over GM potato products, and Monsanto stopped marketing GM potatoes in 2001. An account of mechanisms of transgenic virus resistance can be found in Slater et al. (2008).

### *Amflora in Europe*

There has been considerable resistance to GM potatoes in the European Union. The first GM potato for non food use to be cultivated in Europe was expected to be Amflora. It was a pure amylopectin starch potato developed by BASF Plant Science by switching off the gene for Granule Bound Starch Synthase (GBSS), the key enzyme for the synthesis of amylose. However, in 2009 there were delays in its approval for commercial production because of concerns over the antibiotic resistance marker which it contains. Although the EU Commission did then give authorization, Amflora GM potato was banned by the EU General Court in 2013. However, BASF had already stopped selling Amflora potatoes in the EU because it was struggling to gain a market share amid widespread popular and political resistance. BASF had also developed a late blight (*Phytophthora infestans*) resistant

potato, Fortuna, by using genetic transformation to introduce two resistance genes from the wild tuber-bearing Mexican species *Solanum bulbocastanum*. BASF applied on 31 October 2011 for EU approval for Fortuna for both cultivation and for food and feed use, but announced on 16 January 2012 that it was stopping all of its GM crop developments targeted for the European market due to lack of acceptance of GM technology in Europe. Early in 2013 BASF further announced that the company was no longer seeking to introduce Fortuna onto the European market. BASF Plant Science relocated from Germany to the USA.

### **Late Blight Resistant Potatoes**

Late blight of potatoes, caused by the oomycete *Phytophthora infestans*, is still one of the potentially most devastating diseases of plants worldwide. It is now 170 years since its arrival in 1845 first caused the devastation of the potato crop in Ireland, elsewhere in NW Europe, and in North America. Resistance genes (*R*-genes) were found in the Mexican species *S. demissum* and utilized from 1908 onwards, but they failed to provide durable resistance. So today in northern Europe it is common for farmers to spray 10–15 times per year to control late blight. There is currently, however, much interest in the *R*-genes found in other wild species as they can now be introduced into commercially successful cultivars by genetic transformation. There are also reasons to think they may prove more durable, as discussed in the next chapter. The BASF programme produced GM potato Fortuna by introducing two resistance genes *Rpi-blb1* and *Rpi-blb2* from *Solanum bulbocastanum* into the European processing cultivar Fontane (Storck et al. 2012). Briefly, the vector construct used for the *Agrobacterium*-mediated transformation combined the two genes under control of their native promoters and terminators, and used imidazoline herbicide resistance as the selection marker. In over 5 years of field testing since 2006, in locations throughout Europe, late blight disease was not observed on Fortuna. Furthermore, disease had not been observed on Fortuna in previous greenhouse tests using a large collection of *Phytophthora infestans* isolates from all over the European Union. One of the resistance genes *Rpi-blb2* is available in potato cultivars Bionica and Toluca as a result of conventional introgression.

Work on the discovery, isolation, cloning and introduction of resistance genes by genetic transformation has continued in Europe at Wageningen University (the DuRPh project) and the Sainsbury Laboratory at the John Innes Centre in Norwich. The Wageningen group have provided the first scientific report on the production and functional evaluation of cisgenic *R*-gene stacking in different potato cultivars (Jo et al. 2014). They used marker-free transformation of the cisgenes *Rpi-vnt1.1* (from *S. venturii* from Argentina) and *Rpi-sto1* (from *S. stoloniferum* from Mexico) in combination with PCR selection. Two hundred stem explants from each of three selected cultivars (Atlantic, Bintje and Potae9) were prepared and co-cultivated with an *A. tumefaciens* strain carrying only the two cisgenic *R*-genes between the T-DNA borders of a binary plasmid (the genetic engineering involved is described

in the paper). From 1515 shoots collected and screened by PCR with *Rpi-vnt1.1* and *Rpi-stol* primers, 27 PCR positive shoots were selected. They came from different explants, indicating that they were independent transformation events. From eight of these (two from Atlantic, five from Bintje and one from Potae9), normal looking plants were produced that were vector backbone-free (vector backbone gene-specific PCR analysis) and in which both genes were functionally expressed (based on *Avrvnt1* and *Avrstol* infiltration). The eight plants were resistant to all five *P. infestans* isolates tested (detached leaf tests). They are now undergoing further evaluation. Allowing 2–3 years of field trials, it takes a total of just 3–4 years to produce cisgenic late blight resistant potato cultivars which can be released for seed tuber multiplication; considerably less than conventional introgression breeding.

The Sainsbury Laboratory completed a set of field trials in 2012 of transgenic potato cultivar Desiree containing the *Rpi-vnt1.1* gene (Jones et al. 2014). The gene was introduced by standard transformation methods using an NPT2 resistance marker conferring kanamycin resistance to transformed cells in tissue culture. Some Ecuadorian strains of *P. infestans* can overcome resistance conferred by *Rpi-vnt1.1*, but such races are not present in Europe. Although the trials were done over 3 years in Norwich, England, it was 2012 when the weather was particularly favourable for late blight. The field trial had six replicate plots, each plot comprising three subplots of 16 plants, of which two were transgenic Desiree and one was untransformed Desiree. The first symptoms of late blight from natural infection by *P. infestans* genotype 6\_A1 appeared on 13 July and disease severity was scored on 10 August. The *Rpi-vnt1.1* transgenic plants were fully resistant to late blight and remained so until harvested in the first week of October, whereas non-transgenic Desiree plants had 100 % infected tissue on 10 August. The next step in the project is to identify two other resistance genes which can be added to GM potatoes to give triple protection. This will be done in collaboration with the J.R. Simplot Company in the USA.

### ***Simplot's Programme in the USA***

Innate potatoes produced by the J.R. Simplot Company in the USA are now almost certain to be the first GM potatoes used for human consumption, although it is not yet clear if they will be acceptable to companies such as McDonald's. They were deregulated by the U.S. Department of Agriculture (USDA) on 7 November 2014 and voluntarily reviewed by FDA who concluded on 20 March 2015 that they were as safe and nutritious as their conventional counterparts. They will be licensed to select partners for market testing in 2015. The Simplot programme took 14 years of scientific development, safety assessments, and extensive field tests. It used marker-free transformants and intragenic approaches to silence genes by RNA interference that are key to the expression of black spot bruise, asparagine, and reducing sugars in tubers. In RNA interference double-stranded RNA triggers the degradation of

gene transcripts. The method is intragenic because the inserted genes came from cultivated potatoes (*S. tuberosum*) or a sexually-compatible wild relative (*S. verrucosum* for *Ppo5*). No foreign genes, no antibiotic resistance markers and no vector backbone sequences were incorporated into the plant genome. A plasmid vector called pSIM1278 was used to separately incorporate two silencing “cassettes” into cultivars Ranger Russet, Russet Burbank and Atlantic and into two proprietary chipping (crisping) cultivars. Expression of the first cassette lowered transcript levels for the *Asn1* (asparagine synthetase-1) (Rommens et al. 2008) and *Ppo5* (polyphenol oxidase-5) genes and consequently limited the formation of the acrylamide precursor asparagine, and the formation of impact-induced black spot bruise that occurs when the enzyme polyphenol oxidase oxidizes phenols to produce dark pigments. This browning also normally occurs when untransformed potatoes are cut and left in the air. The presence of black spot bruise results in lower quality and subsequent production losses during processing into French fries or chips (crisps). Asparagine is converted into acrylamide, a chemical compound that has been linked to cancer, when potatoes are fried at high temperatures. Expression of the second cassette reduced the tuber formation of reducing sugars (glucose and fructose) from starch by lowering transcript levels for genes *PhL* (amyloplast-targeted phosphorylase-L) and *R1* (water dikinase) (Rommens et al. 2006). Benefits included improved quality by contributing to the desired golden brown colours required by most French fry and chip customers. Lower reducing sugars in stored potatoes also contribute to lower acrylamide production because they react with the amino acid asparagine in the Maillard reaction to produce acrylamide. As mentioned above, Simplot is currently working with the John Innes Centre in the UK to add transgenic late blight resistance to Innate™ potatoes.

### ***Cultivation of GM Potatoes in Centre of Origin and Diversity***

The cultivation of GM cultivars in the centre of origin and diversity of the potato has long been strongly opposed, partly based on cultural values and partly because of concerns that transgene flow into landraces would alter the natural gene pool and diminish biodiversity. Numerous studies have demonstrated that gene flow can occur in potato between commercial potato cultivars and native landraces, and even wild species (Scurrah et al. 2008), but the fate of foreign genes and their impact on biodiversity is not so clear. Ghislain et al. (2014) have shown that Andean farmers can grow a modern cultivar in the same field as their native landraces and maintain the integrity of both. They screened 1671 DNA samples from more than 400 native cultivars (landraces) from three regions in the Peruvian Andes where the commercial, fertile cultivar Yungay and landraces have coexisted for 15–25 years. Simple sequence repeat (SSR) markers were used to identify putative hybrids based on allele sharing with those of Yungay. None of the landraces were hybrids derived from a successful gene-flow event between ‘Yungay’ and a native potato.

The authors concluded that even when all conditions for gene flow are met, farmers have not unintentionally incorporated genes from the commercial variety 'Yungay' into their native potato germplasm. These results will probably not allay the fears of those opposed to GM crops, but they do provide some evidence for the view that Andean farmers could grow transgenic potatoes among their landraces without destroying the integrity of those landraces.

## Low Lignin Alfalfa

Another GM crop to be deregulated by USDA at the end of 2014 (8 November 2014) was low lignin alfalfa (event KK179) to be marketed as HarvXtra<sup>TM</sup>, a stack with Roundup Ready<sup>®</sup> GM alfalfa which has been grown in the USA since 2005. Lignins are a group of heterogeneous phenylpropane polymers which form important constituents of cell walls within supporting and conducting tissues (xylem cells). The monomeric constituents of lignins are three hydroxycinnamyl alcohols (monolignols): coniferyl alcohol (G lignin),  $\rho$ -coumaryl alcohol (H lignin) and sinapyl alcohol (S lignin). The amounts of lignin in HarvXtra<sup>TM</sup> are generally similar to those found in conventional forage harvested several days earlier under similar production conditions. The reduced lignin alfalfa therefore increases forage quality compared to conventional forage of the same age, maximizes forage yield by delaying harvest for several days, and gives farmers more flexibility in forage harvest timing (James 2014). Details of the genetic engineering can be found in the published Plant Risk Assessment for KK179. Briefly, KK179 reduces lignin through the suppression of caffeoyl-CoA 3-*O*-methyltransferase (CCOMT), a key enzyme in the G lignin biosynthetic pathway that converts caffeoyl-CoA to feruloyl-CoA (Vanholme et al. 2010). KK179 was produced by *Agrobacterium tumefaciens*-mediated transformation of conventional alfalfa R2336 with *CCOMT* gene segments, derived from alfalfa, assembled to form an inverted repeat DNA sequence. The inverted repeat sequence produces double-stranded RNA (dsRNA) which suppresses endogenous *CCOMT* gene expression via the RNA interference (RNAi) pathway. Suppression of the *CCOMT* gene expression leads to lower *CCOMT* protein expression resulting in reduced synthesis of G lignin subunit compared to conventional alfalfa at the same stage of growth. The reduction in G lignin subunit synthesis leads to reduced accumulation of total lignin. The *CCOMT* gene segments are under the control of the *Pal2* promoter from the phenylalanine ammonia-lyase gene in bean (*Phaseolus vulgaris*). The plasmid vector used for transformation had two separate T-DNAs (T-DNA I and T-DNA II), the second of which contained a marker gene for *in vitro* selection of transformed plantlets. The transformed plant was crossed with MS 208, an elite, male sterile, conventional alfalfa plant, in which the unlinked insertions for T-DNA I and T-DNA II were segregated to produce KK179 with confirmation by PCR analysis. The segregation of the two T-DNAs was used to isolate a subset of transformed plants that contained the *CCOMT* suppression cassette (T-DNA I) but did not contain the marker

expression cassette (T-DNA II). This resulted in the subsequent identification of a single marker-free plant line of KK179.

Low-lignin alfalfa cultivars (HiGest® 360 and HiGest® 660) from conventional breeding will also be available on a limited basis in 2015 from Alforex Seeds which is a Dow AgroSciences' Company (<http://www.alforexseeds.com/agronomy/low-lignin-alfalfa>). Alforex breeders surveyed their alfalfa germplasm collection for the various genetic factors that contribute to forage quality and found enough natural genetic variability within the pool to allow significant genetic improvement to be made. Using conventional breeding techniques and screening hundreds of thousands of plants, the breeders selected elite parental material that expressed both strong agronomic traits and low lignin content. The genetic reduction in lignin results in improved fibre digestibility which should increase animal intake, which in turn should increase milk or meat production. Hence once again we see a complex trait that can be improved by both conventional breeding and genetic transformation. However, our final example is one where genetic transformation is required because it is necessary to introduce novel biochemistry.

## Golden Rice

Rice is staple food for approximately half of the world's population, and in favourable conditions it is supplemented with vegetables, fruit, meat and fish to make a nutritionally well-balanced diet. However, these supplements are missing in part or as a whole for hundreds of millions of poverty-stricken people in Asia, Africa and Latin America. One of the most conspicuous signs is caused by the absence of provitamin A, a mixture of  $\beta$ -carotene and chemically related precursors of vitamin A. The human body cannot produce these compounds by itself, but it can convert them to vitamin A. Lack of provitamin A leads to poor eyesight, or even complete blindness, and also weakens the immune defence system (Hahlbrock 2009).

Grains of (milled) rice do not contain any provitamin A that could be increased by conventional breeding, but an understanding of carotenoid biosynthesis resulted in rice being genetically engineered by research workers in Switzerland to produce provitamin A. Because of the rice's new golden-yellow colour it became known as Golden Rice (Fig. 17.1). An account of the genetic engineering can be found in Slater et al. (2008).

Briefly, immature rice endosperm is capable of synthesizing geranylgeranyl diphosphate (GGDP), which is also known as geranylgeranyl pyrophosphate. The head to head addition of two molecules of GGDP to form phytoene is catalyzed by phytoene synthase, and is the first dedicated step in the formation of carotenoids (Fig. 17.2). Structurally carotenoids are terpenoids, derived by condensation of prenyl pyrophosphates, whose conjugated double-bond system determines their colour and is responsible for their biological actions as antioxidants. Initial studies of the carotenoid biosynthetic pathway indicated that production of Golden Rice would require three genes encoding a phytoene synthase, a carotene desaturase

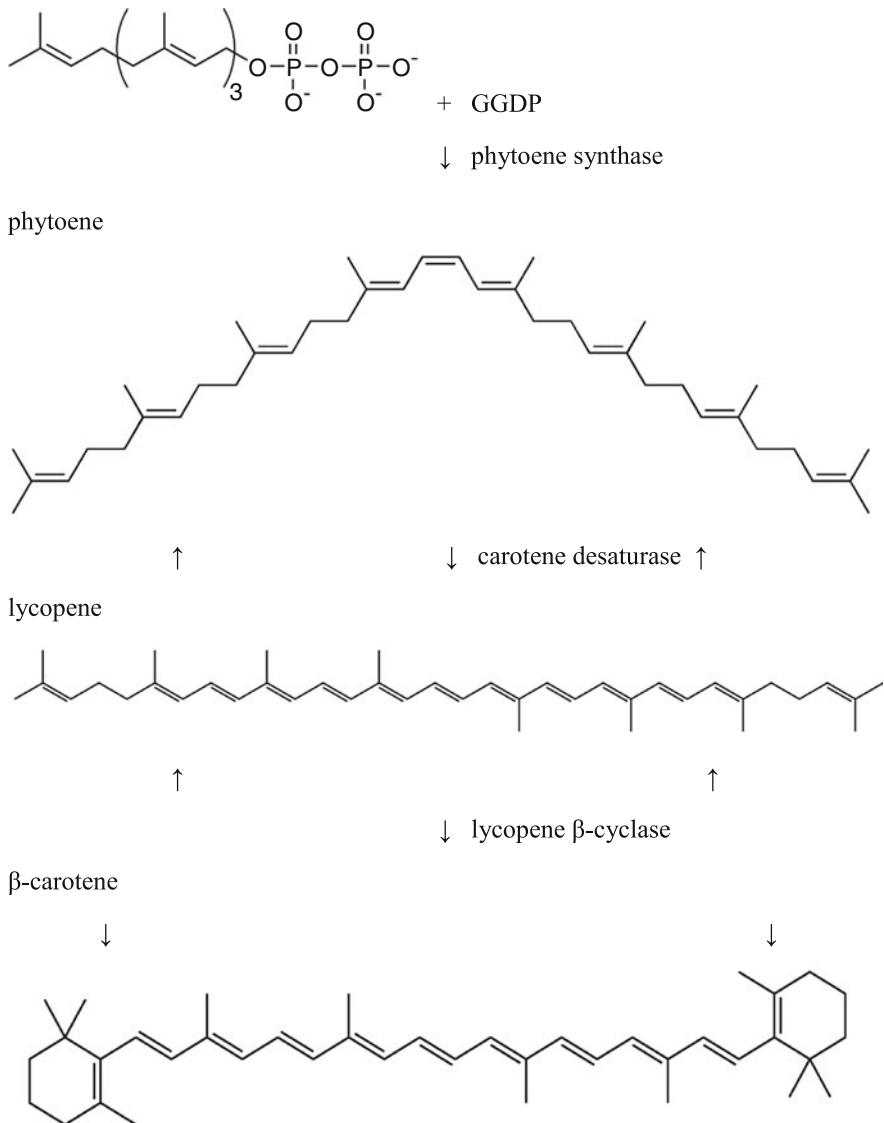


**Fig. 17.1** Bowls of white and golden rice (source: Wikipedia “Golden Rice” by International Rice Research Institute (IRRI). <http://www.flickr.com/photos/ricephotos/5516789000/in/set-72157626241604366>. Licensed under CC BY 2.0 via Commons; [https://commons.wikimedia.org/wiki/File:Golden\\_Rice.jpg#/media/File:Golden\\_Rice.jpg](https://commons.wikimedia.org/wiki/File:Golden_Rice.jpg#/media/File:Golden_Rice.jpg)

capable of introducing four double bonds to form lycopene, and a lycopene  $\beta$ -cyclase to form the rings in  $\beta$ -carotene. In 2000, Ye et al. (2000) transformed japonica rice with three transgenes (from daffodil, a bacterium and daffodil, respectively) plus a selectable marker gene by *Agrobacterium* co-transformation of two constructs that produced enzymes targeted to the plastid, the site of GGDP synthesis. The highest-producing line secured was found to contain  $1.6 \mu\text{g} \beta\text{-carotene g}^{-1}$  endosperm. This level was calculated to make a significant contribution to the daily intake of vitamin A but was not sufficient to provide the complete dietary requirement of the vitamin.

Further work was done as a public-private partnership with Syngenta to help further develop Golden Rice. Paine et al. (2005) hypothesized that the daffodil gene encoding phytoene synthase (*psy*) was the limiting step in  $\beta$ -carotene accumulation. Through systematic testing of other plant *psy*'s, they identified one from maize that substantially increased carotenoid accumulation in a model plant system. They then went on to develop ‘Golden Rice 2’ (GR2) by introducing into indica rice, the maize *psy* gene in combination with the *Erwinia uredovora* carotene desaturase (*crtl*) gene used to generate the original Golden Rice. The construct used for transformation pSYN12424 contained the *crtl* and maize phytoene synthase (*ZmPSY*) genes both driven by the endosperm-specific rice glutelin promoter (*Gt1p*), and the phosphomannose isomerase (*pmi*) selectable marker gene (to avoid antibiotic selection) fused to the ubiquitin gene promoter (*ubilp*). The lycopene  $\beta$ -cyclase transgene proved unnecessary to produce  $\beta$ -carotene

geranylgeranyl pyrophosphate also known as  
geranylgeranyl diphosphate (GGDP)



**Fig. 17.2** Part of carotenoid biosynthetic pathway showing carbon skeletons

in the endosperm. The outcome was an increase in total carotenoids of up to 23-fold (maximum 37  $\mu\text{g/g}$ ) compared to the original Golden Rice and a preferential accumulation of  $\beta$ -carotene (about 30  $\mu\text{g } \beta\text{-carotene g}^{-1}$  endosperm). This level

was considered probably adequate for provitamins ingested with food to be nutritionally effective.

Golden Rice was initially developed in public laboratories with public funding for humanitarian purposes. The researchers wanted to patent-protect the process to fulfil the goal of supplying it to subsistence farmers free of charge and restriction. But there was a problem. The research underpinning the development of Golden Rice had involved the use of procedures and technology protected by 70 Intellectual Property Rights and technology protection rights belonging to 32 different companies and universities (Slater et al. 2008). Eventually, all participating seed companies and patent holders agreed to forgo profits as long as the annual income of recipient small farmers remained below US\$10,000. However, there was still some way to go as it was necessary to transfer the new trait into locally adapted cultivars for use in developing countries, either by breeding or new transformations of elite cultivars. It was also necessary to overcome traditional habits of eating white rice (Hahlbrock 2009). Golden Rice is still being developed by the Philippine Rice Research Institute (PhilRice) and the International Rice Research Institute (IRRI). According to James (2014), IRRI reported in March 2014 that the research, analysis and testing of  $\beta$ -carotene enriched Golden Rice (GR2) was continuing in partnership with collaborating national research agencies in the Philippines, Indonesia and Bangladesh. In multi-locational trials to date, the target level of  $\beta$ -carotene in the grain has been achieved but yields have on average been lower than those from comparable local cultivars already preferred by farmers. Therefore at IRRI the Golden Rice trait (GR2) is being bred into high yielding cultivars to get suitable advanced lines for new trialling.

# **Chapter 18**

## **Climate Change and Resistance to Pests and Diseases**

### **Introduction**

Feeding the world's expanding population is a tremendous challenge, particularly during a period of climate change. Substantial progress will be required in bridging the yield gap between what is currently achieved per unit of land and what should be possible in future, with the best farming methods and best storage and transportation of food, given the availability of suitably adapted cultivars. The CO<sub>2</sub> concentration in the earth's atmosphere is predicted to increase from 370 ppm in 2015 to 550 ppm in 2050. As this is a global phenomenon, all plant breeding with field crops will take place in an atmosphere of increasing CO<sub>2</sub> levels and plant breeders will be selecting new cultivars that are better adapted to a CO<sub>2</sub>-enriched atmosphere. It remains to be seen if the resulting cultivars have higher yields than those being grown today. The situation is complicated by the fact that other environmental changes affecting crop production are predicted, only some of which are directly related to climate change. The gaseous changes in the atmosphere are predicted to change the earth's climate, making it warmer by an average of about 2 °C, even if effective steps are taken to reduce greenhouse gas emissions. The predicted rise in global temperatures and changes in patterns of precipitation will affect different regions in different ways. Hence various changes in farming systems in different parts of the world can be anticipated in terms of crops grown and growing seasons. Breeders will need to select new cultivars better adapted to these new environmental conditions. The land area used for agriculture may remain static or even decrease as a result of degradation and climate change unless crops can be bred which tolerate the abiotic stresses associated with areas considered marginal for growing crops. These can be due to drought, high temperatures, low soil fertility, salinity or aluminium toxicity in acid soils. Salinity is also a major threat to the sustainable irrigation required to maintain the yield and quality of many crops. As we saw in the chapter on Genotype × Environment interactions, the challenge for breeders is to combine tolerance to abiotic stresses with high yield potential in more favourable

environments; unless breeders are going to produce a large number of locally adapted cultivars. Returning to irrigation, there is going to be increased competition for water resources between agricultural, industrial and domestic users. The most pressing constraint on food production is likely to be water scarcity, despite some increases in precipitation in some areas. Hence there is a need to select cultivars for more efficient use of water if environmentally sustainable increases in crop production are to be achieved. Likewise there is a need to increase the efficiency of uptake and assimilation of nutrients into dry matter yield and protein, as also considered in the chapter on Genotype  $\times$  Environment interactions. Crops often fail to reach their yield potential because of competition from weeds and losses due to attacks by pests and diseases. Cultivars with inbuilt resistance to pests and diseases, as well as the ability to suppress weeds, would address all of these concerns. Hence these will remain highly desirable breeding goals for the indefinite future. Priorities may change as a result of climate change because pathogens and pests have specific temperature and moisture requirements for their population growth. This leads us into the last topic for consideration in this book, breeding for resistance to pests and diseases. It differs from breeding for other aspects of the crop environment because we are dealing with interactions between plants and other living organisms with the capacity to evolve by natural selection.

## Breeding for Resistance to Pests and Diseases

Serious disease problems occur in a crop when susceptible cultivars are attacked by virulent and aggressive pathogens in environmental conditions favourable to disease development and chemical control is unavailable, ineffective or too expensive. In extreme circumstances the complete failure of a staple crop can result in a famine with serious political and societal consequences, as occurred with late blight of potato in Ireland in 1845, 1846, 1848 and 1849 (Nally 2011) (Fig. 18.1). Increased levels of resistance to major diseases are therefore desirable, and even when chemical control is effective and affordable, inbuilt resistance will reduce production costs. The same will be true of pests that feed off the crop, some of which may also transmit diseases. It is sometimes useful to distinguish between pathogen and parasite. A parasite lives in or on another organism from which it obtains nutrient, whereas a pathogen is a parasite that produces a disease in its host (Day 1974). Obligate parasites cannot multiply in nature without their hosts whereas facultative ones can grow and live on substrates other than a living host. The breeding goal will be to combine adequate levels of resistance with the high yields and quality required for a new cultivar to be adopted by growers and end users. Adequate resistance means sufficient to prevent a serious disease or pest outbreak, but some resistance is better than none, particularly as part of integrated disease and pest management which also involves some chemical and cultural control. Furthermore, the breeder will hope that the resistance proves durable and remains effective over a wide area of cultivation for many years. The problem is that durability of resistance

**Fig. 18.1** One consequence of the Irish potato famine of 1845 and 1846. (a) Grosse Île: Canadian Quarantine Station for Immigrants from 1832 to 1937. (b) 1847: 100,000 immigrants, mainly Irish; over 5000 died of typhus and other diseases. (c) Irish memorial on Grosse Île, erected in 1909



is not a property of the new cultivar *per se*. It depends on the population biology and genetics of the pathogen (or pest) and the epidemiology of the disease (or pest). It also depends on whether or not resistance genes can be effectively managed in the environment. Durability is still difficult to predict with certainty and there have been mixed fortunes in the past. Experience has shown that some sources of resistance are not durable and hence best avoided, a well known example being the *R*-genes from *Solanum demissum* for resistance to late blight of potato caused by the oomycete pathogen *Phytophthora infestans*, which we will be considering later in this chapter. In thinking about novel strategies and sources of resistance it is always worth bearing in mind that a particular crop will be a non host for most of the diseases and pests of other crops.

The prerequisites for successful breeding are a source of potentially durable resistance, a reliable screen for resistance using the most appropriate isolate(s) or strain(s) of the parasite and an understanding of the inheritance of resistance. However, a breeder must be realistic in deciding priorities and not take on so many objectives that little progress can be expected for each one. This means that for less important pests and diseases, potential cultivars should simply be screened to avoid extreme susceptibility. This can be done in tests, or when natural epidemics of minor pests and diseases occur. For potatoes globally, Bradshaw and Bonierbale (2010) list 7 major pests (a moth, a beetle, a weevil, an aphid, a fly and 2 types of nematode) and 25 major diseases (1 oomycete, 11 fungal, 1 actinomycete, 3 bacterial and 9 viral or viroid diseases). The breeding programme at the Scottish Plant Breeding Station and the Scottish Crop Research Institute from 1920 to 2008 concentrated on breeding for resistance to late blight (the oomycete), three viruses (*PVY*, *PVX* and *Potato leafroll virus*) and two cyst nematodes (*Globodera rostochiensis* and *G. pallida*), and involved introgression of resistance genes from the wild and cultivated species of Latin America (Bradshaw 2009a). For other problem-causing diseases (fungal, bacterial and viral) in the UK and export markets, clones in intermediate and/or advanced trials were assessed for their level of resistance to discard any with extreme susceptibility. High levels of resistance to these other diseases in new cultivars occurred largely by chance and susceptibility was accepted if the cultivar had other desirable traits. Bradshaw (2009a) concluded that this situation was unlikely to change unless end users demanded higher levels of resistance (for example, to blemish diseases) because they could see economic advantages, or governments legislated for it because they wanted environmental benefits (for example, resistance to common scab, the actinomycete, as a result of pressure to reduce the use of water for irrigation). In the meantime in the UK control measures other than resistant cultivars are required for many pest and disease problems of potatoes. We will now focus on the principles of breeding for durable resistance, with appropriate examples of successes and failures, often from potatoes, rather than attempt a comprehensive review. *Breeding Major Food Staples* (Kang and Priyadarshan 2007) can be consulted for information on the world's major food staples and also the *Handbook of Plant Breeding* (Prohens et al. 2008) for these and other crops.

## Types of Resistance

In seeking sources of resistance, breeders need to understand the evolutionary history of their crop, its genetic base, and its relationships with its wild relatives. It is easiest to use resistances already present in cultivars, but can they provide adequate levels of resistance? If not, is resistance available in landraces or in wild relatives? And what type of resistance is most likely to combine adequate control of the pest or disease with durability?

I find that the most useful classifications of resistance are qualitative versus quantitative and specific versus broad spectrum or general resistance. Others will no doubt have their own preferences and further terminology will be introduced as we progress. Qualitative resistance is where plants from a cross can be classified unambiguously as resistant or susceptible and the number in each category can be counted. The difference between resistant and susceptible plants is usually large with resistant ones free, or showing just traces of disease, and susceptible ones showing extensive symptoms (infection). Mendelian ratios and the corresponding genes can be sought. Qualitative resistance has proved to be oligogenic, controlled by one or a few major genes, as we shall see later in this chapter. In contrast, quantitative resistance is where there are no unambiguous categories, rather degrees of resistance or susceptibility which are measured on some scale. Genes are identified indirectly through their associations with molecular markers and have proved to be few or many in number. Hence it seems reasonable to talk about qualitative oligogenic resistance versus quantitative resistance. Both types of resistance can be assessed on plant parts (e.g. detached leaves), seedlings or whole plants in controlled environment facilities such as growth chambers or specialized glasshouses. Under these circumstances the breeder or pathologist, for example, may apply inoculum of the pathogen containing a known concentration of spores and wait to see how many lesions develop, how quickly they spread, how soon and for how long they sporulate and how many spores they produce. Some disease tests will give clear cut qualitative differences: plants or detached leaves are either susceptible, producing sporulating lesions, or resistant, showing complete immunity to infection or a hypersensitive reaction in which infected cells die. Other tests will give quantitative results, showing variation in the number of lesions, their spread, and spore production. Both types of resistance can also be assessed on plants in the field, often in special nurseries, where the progress of the disease can be followed under more natural but less controlled conditions. Hence care is needed in using the term field resistance as it need not necessarily imply quantitative resistance, although often it will. Disease progress is usually the result of a number of cycles of infection and sporulation, with the outcome referred to as an epidemic. We will look at measures of disease progress in a later section. Similar considerations apply to pest resistance.

The pathogen (or pest) may be differentiated into species, subspecies, biovars or physiological races; examples of all of these can be found among the many pests and diseases of potato (Bradshaw et al. 2000). A cultivar (or potential cultivar) may

be resistant to some isolates of a pathogen but susceptible to other isolates (isolate specific) or it may be resistant to most or all known isolates of a pathogen (broad spectrum). A cultivar may be more or less susceptible to different isolates of pathogen, the isolates thus varying in their aggressiveness. Once many potential cultivars are considered, cultivar  $\times$  isolate interactions may occur which are relevant to the programme's objectives, and which determine the most appropriate isolate or isolates of the pathogen for use in screening. The results are usually clearer when both the potential cultivars and pathogen isolates are genetically homogeneous.

## **Qualitative Oligogenic Resistance: The Gene-for-Gene Concept**

Plant breeders frequently thought that they had made breakthroughs in breeding for disease resistance during the first half of the twentieth century, only to be disappointed. They found simply inherited sources of high levels of resistance in landraces and wild relatives that they introgressed into new cultivars, but after just a few years of use the resistant crops often succumbed once more to disease; so called boom and bust cycles. The start of genetic studies and scientific breeding for resistance was Rowland Biffen's demonstration that resistance to yellow rust (*Puccinia striiformis*) in wheat obeyed Mendel's laws: the resistance of Rivet wheat was determined by a single recessive gene (Biffen 1905, 1912). The appeal of simply inherited qualitative resistance to breeders, geneticists and pathologists can be seen in a 1971 literature review by Person and Sidhu (1971): out of 912 papers published since 1912 on the genetics of plant disease resistance, 875 reported resistances due to major genes and just 60 reported quantitative resistances. The appeal to molecular geneticists and pathologists is just as great today for understanding the genetics of host-parasite interactions. Let us therefore start with the gene-for-gene hypothesis before considering one of the big disappointments.

### ***Resistance to the Rust Melampsora lini in Flax: Flor's Gene-for-Gene Concept***

The gene-for-gene concept was put forward in 1942 by Harold H Flor who was working on resistance to the rust *Melampsora lini* in flax (*Linum usitatissimum*) and did genetic studies on both host and pathogen (Flor 1942). A summary of this and subsequent research by Flor can be found in the book by Day (1974), which also covers genetic systems in fungi, and explains the different types of spores produced by asexual and sexual reproduction. Day (1974) was also able to give a number of

examples of where the gene-for-gene concept had either been demonstrated or suggested for host-parasite systems including rusts, smuts, bunts, mildews, other parasitic fungi, nematodes, insects, bacteria, viruses and higher plant parasites of economically important crops.

Returning to flax, Flor found that cultivar Bombay was resistant to race 22 but susceptible to race 24 of *M. lini*. In contrast, cultivar Ottawa was resistant to race 24 but susceptible to race 22. The F<sub>1</sub> between Bombay and Ottawa was resistant to both races but in the F<sub>2</sub> there were four phenotypic classes, resistant to both races, resistant to race 22, resistant to race 24 and resistant to neither, in ratios not significantly different from 9:3:3:1. In other words there was independent assortment of two dominant genes, one conferring resistance to race 22 and the other resistance to race 24. When the two races of *M. lini* were crossed, their 'F<sub>1</sub>' was avirulent on both cultivars (they were resistant) but in the 'F<sub>2</sub>' there were four phenotypic classes of cultures (dikaryons), avirulent on both cultivars, avirulent on Bombay, avirulent on Ottawa and avirulent on neither (i.e. virulent on both), in ratios not significantly different from 9:3:3:1. In other words there was independent assortment of two dominant genes, one avirulent on Bombay and the other on Ottawa. The resistant outcome was the result of an interaction between the products of the resistant allele in the host and the avirulent allele in the pathogen. The pathogen becomes pathogenic when it fails to produce the avirulent product and is not recognized as a pathogen by the host with the *R* gene. Thus if *R* and *r* are the resistant and susceptible alleles at a locus in the host, and *Avr* and *avr* the avirulent and virulent alleles at a locus in the pathogen, we have:

		Pathogen	
		<i>Avr</i>	<i>avr</i>
Host	<i>R</i>	R	S
	<i>r</i>	S	S

where R is an incompatible interaction resulting in resistance and S is a compatible one resulting in susceptibility. In today's terminology we say that the *R* gene encodes a receptor for the pathogen ligand (effector protein) specified by the *Avr* gene.

Flor found that genes for resistance from different sources that mapped to the same locus could often be distinguished by their reactions with certain rust races (Day 1974). The different genes were assumed to be multiple alleles at one locus. Flor was able to assign 26 different resistance genes to five independent loci: *K* (1 allele), *L* (12), *M* (6), *N* (3) and *P* (4). Furthermore, different alleles at the same locus recombine at low frequencies during meiosis to produce new 'alleles'. Rust resistant flax cultivars generally carried only one resistance gene, but Flor and Comstock (1971) developed lines with three different resistance genes at three loci. They independently backcrossed each gene into cultivar Bison to produce

monogenic Bison-like lines, then intercrossed them to combine resistance genes in pairs, and finally intercrossed these to combine three genes in a single line; an example of pyramiding genes in the hope of increasing their durability.

After this early work by Flor, extensive research was done on the flax *L* locus which was found to be a single-copy gene that can encode TIR-NBS-LRR proteins with at least 13 distinct recognition specificities for different rust effector proteins. It became a model system for understanding how the TIR-NBS-LRR proteins recognize the effector proteins, become active and trigger a defence response that results in resistance through host cell death. Ellis et al. (1999) sequenced the 13 known *L* alleles and compared them to gain insight into their evolution and resistance specificity. The LRR region appeared to be the main determinant of specificity differences between alleles, and functional analysis in transgenic plants of recombinant alleles confirmed the LRR specificity of *L*2. However, specificity differences were also found in the TIR region. The sequence comparisons also indicated that the evolution of *L* alleles had probably involved reassortment by intragenic recombination of variation from accumulated point mutations, as well as deletion and duplication events in the LRR-encoding region. More recently Bernoux et al. (2011) have shown that the TIR domain of the L6 protein is dispensable for recognition of the AvrL567 effector but is required for induction of the defence response. The resistance protein activation was thought to result from exposure of the TIR signalling domain following a conformational change of the NB-ARC domain upon effector recognition.

### ***Examples from Other Crop-Parasite Interactions***

Jones (2001) reviewed the progress that had been made during the 1990s in isolating *R* genes from other well characterized ‘gene-for-gene’ resistance in plants and studying their evolution and mechanisms of resistance. He covered the *Rpl* (*Resistance to Puccinia sorghi* 1) locus for rust resistance in maize with at least 14 recognition specificities; the *Mla* (*Mildew resistance a*) locus for powdery mildew resistance in barley with 28 *Mla* alleles; the *Cf-9* (*resistance to Cladosporium fulvum*-9) and orthologous *Cf-4* locus of tomato; the *Mi* gene for resistance to the nematode *Meloidogyne incognita* in tomato (*Mi-1.2*); and the *Xa21* gene for *Xanthomonas* resistance in rice, as well as the *R*-gene complement of *Arabidopsis*. Of particular significance to plant breeders, Jones pointed out that some, though not all, mutations in *Avr* genes lead to reduced pathogen fitness, and that there was also evidence that *R* genes could lead to reduced plant fitness in the absence of a pathogen. Hence the stacking (pyramiding) of *R*-gene alleles from different loci in a single cultivar might not provide durable resistance, will select physiological races that can tolerate the loss of multiple compatibility factors, and may lead to yield penalties. We will shortly encounter such an example that also emphasizes the point that durability is difficult to predict. It is still not clear if *R* genes can be found which recognize indispensable effectors (*Avr*'s) and hence

will be durable *per se*. In Chap. 13 we encountered cultivar mixtures as a way of reducing disease incidence. The isolation (cloning) of *R* genes has opened up a new possibility, true multilines, in which genetic transformation is used to produce lines or clones of a successful cultivar with different *R*-gene alleles (from a single locus or different loci). These can then be grown as a mixture in which all components are genetically identical to each other and to the cultivar which was transformed, apart from carrying different resistance alleles. The cultivar should therefore still be commercially acceptable but have a useful level of durable resistance, provided there is a cost to virulence so that the mixed-pathogen population never completely loses the corresponding *Avr* genes (Jones 2001).

Jones et al. (2014) have provided a more recent review of the interaction between effector *Avr*-genes and *R*-genes, with late blight of potatoes (*Phytophthora infestans*) as an example (*R*-genes are designated *Rpi*-genes). Hosts and pathogens have co-evolved large *R*-gene and effector repertoires, respectively; potato and *P. infestans* being no exception. The sequenced doubled-monoploid potato clone DM and *P. infestans* have more than 438 *R*-genes and 563 effector candidates, respectively. Jones et al. (2014) explain that some *R*-genes are more durable than others because some effectors are more indispensable to the pathogen than others. Effector molecules are used by pathogens to suppress host defence mechanisms. They are usually delivered into host cells but can act in the intercellular spaces of the leaf. Effectors interfere with defence processes directly, or with their activation, by disrupting host-signalling mechanisms. However, plants have evolved the capacity to recognize specific effectors, either directly or indirectly by detecting their effects on host components. This recognition is usually mediated by intracellular receptors encoded by disease resistance (*R*) genes. Advances in DNA sequencing methods means that the effector repertoire of pathogens can be analyzed and effectors predicted from their protein sequence motifs. For example, *P. infestans* effectors have a signal peptide and a so-called RxLR amino acid sequence motif close to the signal peptide cleavage site in the secreted effector. If a particular effector is shared between multiple races of a pathogen, then it is more likely to be indispensable for the pathogen than an effector that is present in some but not all races. This knowledge provides a crucial tool for discovering the most indispensable pathogen effectors, and prioritizing *Rpi*-genes that recognize these effectors for deployment in crops. The importance of the *P. infestans* *Avr3a* effector was directly examined by testing the virulence of races in which the *Avr3a* gene had been silenced. Remarkably, *P. infestans* lines with reduced levels of *Avr3a* expression are either weakly virulent or completely non-virulent. In other words, an effector needs to be produced for virulence. *Avr3a* is found in two allelic forms, *Avr3a<sup>KI</sup>* and *Avr3a<sup>EM</sup>* (varying in two amino acid positions). The *Rpi* gene *R3a* confers recognition of the *Avr3a<sup>KI</sup>* form but not the *Avr3a<sup>EM</sup>* form. The *Avr3a<sup>EM</sup>* form is present in the widespread virulent race 13\_A2, which is why it can overcome *R3a*. Efforts are underway in at least two laboratories to identify novel forms of *R3a* that can recognize both *Avr3a<sup>KI</sup>* and *Avr3a<sup>EM</sup>*, for example by screening wild *Solanum* populations followed by transient *Agrobacterium*-mediated assays of *R3a* and *Avr3a<sup>EM</sup>* alleles.

Recently Du et al. (2015) demonstrated that the receptor-like protein ELR (elicitin response) from the wild potato *Solanum microdontum* mediates extracellular recognition of the elicitin domain, a molecular pattern that is conserved in *Phytophthora* species. ELR associates with the immune co-receptor BAK1/SERK3 and mediates broad-spectrum recognition of elicitin proteins from several *Phytophthora* species, including four diverse elicitins from *P. infestans*. Transfer of ELR into cultivated potato resulted in enhanced resistance to *P. infestans*. Hence pyramiding cell surface pattern recognition receptors with intracellular immune receptors could maximize the potential of generating a broader and potentially more durable resistance to *P. infestans*.

## ***Physiological Races and Host Differentials***

The gene-for-gene hypothesis predicts the number of physiological races a set of resistant cultivars will differentiate, as explained by Day (1974). With  $n$  genes for resistance, each of which may have two phenotypes (resistant or susceptible), there are  $2^n$  different races which can be identified with the  $n$  monogenic differentials. These need to be agreed and maintained for the benefit of breeders and pathologists working with a particular crop-parasite combination. The scheme developed by Black et al. (1953) for late blight of potatoes was the following. The genes for resistance were numbered  $R1, R2, R3, \dots, R11$  and the races of *Phytophthora infestans* were numbered to indicate their virulence towards them. Race 0 was avirulent on all these genes whereas race 1, 2, 3 was virulent on any combination of the genes  $R1, R2$  and  $R3$ , and so on. The 11 late blight differentials are monogenic for the resistance genes, apart from R9 which unsatisfactorily contains  $R1, R2, R3$  and  $R9$ . Other schemes for race names and race formulae are discussed by Day (1974).

## ***R-Genes from Solanum demissum for Resistance to Late Blight of Potatoes***

The devastation of the potato crop in North America after the arrival of late blight in 1842 or 1843, and in Ireland and NW Europe after its arrival in 1845, led to greater breeding efforts throughout Europe and North America. Although no highly resistant or immune cultivars had been produced by the beginning of the twentieth century, sufficient resistance was supposedly built up to prevent further disasters. The apparent breakthrough was the discovery of the  $R$ -genes for high levels of resistance in *Solanum demissum*. Introgression breeding started in 1908 in Germany and 1909 in Great Britain (Muller and Black 1951). Cultivar Craigs Snow White was released from the Scottish Plant Breeding Station (SPBS) in 1948 from a cross

made in 1939 between Craigs Defiance and clone W800(2). The latter clone was bred by Dr. J. H. Wilson of St Andrews, Scotland, from germplasm including *S. demissum*, and given to the newly founded SPBS after his death in January 1920. Craigs Snow White contained what became known as the *R1* gene and was designated the *R1* differential once appropriate genetic studies had been done. As reviewed by Van der Plank (1968), it failed to provide durable resistance. By 1955 cultivars with the *R1* gene were being grown widely in West Germany and races virulent on *R1* were frequent enough for the gene to delay a blight epidemic by only 5 days. In contrast, in the Netherlands in 1956 there were no commercial cultivars of any importance with the *R1* gene, virulent races were rare, and it delayed blight epidemics by about a month. However, the rate of increase of virulent races could be seen in the USA. Cultivar Kennebec was the first one with the *R1* gene to be grown widely in the USA. When released in 1948 it either escaped blight altogether, or had a few lesions too late in the season to do any harm. By 1954 it had increased in popularity, but virulent races were now abundant, and it was no longer very resistant. Cultivar Pentland Ace (the *R3* differential with gene *R3a*) was released from SPBS in 1951 but soon succumbed to blight. The commercially more successful cultivar Pentland Dell, used for making French fries, entered production in Britain in 1963 when race 4 was the prevalent race of *P. infestans*. It had three genes (*R1*, *R2* and *R3*), each of which conferred resistance to race 4. It succumbed to blight in 1967, but nevertheless went on to become one of SPBS's most successful cultivars and was still the 12th most widely grown cultivar in Britain in 2013, albeit under the protection of fungicides. By 1968 the number of *R*-genes identified had risen to the 11 which are currently recognized, but races of *P. infestans* overcoming the more recently discovered ones (*R5* to *R11*) were already widely distributed in Britain, despite the *R*-genes not being present in common commercial cultivars (Malcolmson 1969). It was clear that they would not provide durable resistance, either singly or in combination, due to the evolution of new races of *P. infestans* which arose by mutation in this asexually reproducing oomycete (only one mating type was present at the time). As a consequence, many breeders started to select for quantitative resistance.

Up until 1984 the epidemics of late blight outside Mexico had been caused by the asexual reproduction of strains of the A1 mating type of *P. infestans*. Yoshida et al. (2013) compared the genomes of 11 herbarium and 15 modern strains and concluded that the original nineteenth century pandemic was caused by a single *P. infestans* lineage, but that this lineage was not the direct ancestor of the one (US-1) that came to dominate the global *P. infestans* population during much of the twentieth century. Rather, a small *P. infestans* metapopulation was established at the periphery of its Mexican centre of origin, or possibly in North America, some time before the first global *P. infestans* pandemic. The first lineage to spread from there was HERB-1, which persisted globally for at least half a century. Subsequently, the US-1 lineage expanded and spread from the metapopulation, replacing HERB-1. An interesting major genomic difference between the HERB-1 and US-1 lineages is the shift in ploidy, from diploid to triploid and even tetraploid. Since 1984, new populations of *P. infestans*, comprising both mating types, have been

spreading from Mexico to the rest of the world (Goodwin and Drenth 1997). Hence there is the possibility of sexual reproduction resulting in oospores which can over-winter in the soil and start epidemics earlier each season. Sexual reproduction may also allow the faster evolution of more virulent and aggressive strains of the pathogen. The new populations of *P. infestans* have prompted fresh screenings of germplasm collections for new sources of resistance and new *R*-genes have and are being found, mapped and cloned, in species other than *S. demissum*, such as *S. bulbocastanum*. Currently, there is much interest and debate over whether or not these *R*-genes will be more durable *per se* or can be deployed in a more durable way (Goverse and Struik 2009).

### ***Successes and Partial Successes***

The *R*-genes from *Solanum demissum* for resistance to late blight of potatoes proved a big disappointment at the Scottish Plant Breeding Station (SPBS) and elsewhere. But there were successes and partial successes with major gene resistance to other pests and diseases (Bradshaw 2009a). Back in 1920, wart (*Synchytrium endobioticum*) was considered a serious, disfiguring and blemish-forming, persistent soil-borne fungal disease of potato tubers. It also provides the first example of success in breeding for disease resistance at SPBS. Among the parents available for use in 1920 was resistance to wart which was inherited in a simple dominant manner (two major dominant genes have now been mapped). Wart was largely eliminated from Britain during the first half of the twentieth century by coupling the breeding of cultivars with immunity to race 1 (the only race found in Britain) with scheduling of land, which prohibited the cultivation of susceptible cultivars on land known to be infested. All SPBS cultivars released from 1934 to 1990 were immune. However, in other European countries wart is still a problem and other races occur. Also available in 1920 were strain-specific resistances to the mosaic viruses, *Potexvirus Potato virus X* (PVX) and its variant *Potato virus B*, and *Potyvirus* members *Potato virus Y* (PVY) strain C and *Potato virus A*. Subsequently, genes conferring comprehensive resistance to all strains of these viruses became available and preferred, although the *Ny<sub>tbr</sub>* (*Tuberosum*) gene for resistance to the common PVY<sup>O</sup> strain of PVY was introduced into the programme in 1948 and proved extremely useful. Thus four genes (two *Ny* and two *Ry*) for comprehensive resistance to PVY, and two genes (both *Rx*) for immunity to all common strains of PVX (resistance-breaking strains have been found), proved of value. The hybrid parent MPI 61.303/34 from the Max Plank Institute in Köln has also been used extensively in European potato breeding as the source of the *Ry<sub>sto</sub>* gene for extreme resistance to PVY. The widely used *H1* gene from an accession (CPC 1673) of Andigena potatoes has remained effective against the golden potato cyst nematode (*Globodera rostochiensis*) in Britain because Ro1 is still the main pathotype, but its widespread deployment has encouraged the spread of the white potato cyst nematode (*G. pallida*). The sources and chromosomal locations of these

genes, along with relevant references, can be found in the book chapter by Bradshaw (2007). In conclusion, viral diseases and soil-borne pests and diseases were controlled much more successfully with major genes than late blight.

## Horizontal and Vertical Resistance

Two idealized situations have given rise to further terminology which can be found in the literature on breeding for resistance, terminology which some authors have used in discussing breeding strategy. We shall see that broad-spectrum quantitative resistance equates with horizontal resistance and race-specific qualitative oligogenic resistance with vertical resistance. The example (Table 18.1) of qualitative resistance shows a host genotype  $\times$  pathogen isolate interaction. As one genotype is susceptible to isolate 1 but two genotypes are susceptible to isolate 2, we say that isolate 2 is more virulent than isolate 1. We also say that genotype 2 displays isolate specific resistance, which is termed race specific if the isolates are physiological races of the pathogen, whereas genotype 1 has broad spectrum resistance, albeit to two isolates in our example. In contrast, there is no host genotype  $\times$  pathogen isolate interaction in the example of quantitative resistance. We say that genotype 4 is more resistant than genotype 5 which in turn is more resistant (less susceptible) than genotype 6, and that isolate 4 is more aggressive than isolate 3. If we just had the results of isolate 3 on genotype 5, we couldn't infer their respective levels of aggressiveness and resistance. The term pathogenicity (ability to cause disease) includes both aggressiveness and virulence. In his book *Disease Resistance in Plants*, Van der Plank (1968) called the resistance of genotype 2 vertical and the resistances of genotypes 4–6 horizontal, and developed what proved to be a controversial but testable theory of vertical and horizontal resistance. In other words, not everyone was convinced of its value. Despite the controversies, J.E. Van der Plank is acknowledged as a founding father of modern quantitative epidemiology.

Van der Plank argued that vertical resistance was resistance to infection whereas horizontal resistance decreased the rate of infection (disease progress). Furthermore, to overcome vertical resistance the pathogen had to become less aggressive on susceptible cultivars of the host, whereas to overcome horizontal resistance the pathogen had to become more aggressive on susceptible cultivars as well. He

**Table 18.1** Qualitative and quantitative resistance

Qualitative			Quantitative		
Host	Pathogen		Host	Pathogen	
	Isolate 1	Isolate 2		Isolate 3	Isolate 4
Genotype 1	Resistant	Resistant	Genotype 4	10 % disease	30 % disease
Genotype 2	Resistant	Susceptible	Genotype 5	30 % disease	50 % disease
Genotype 3	Susceptible	Susceptible	Genotype 6	50 % disease	70 % disease

concluded that when the most virulent races are not the fittest to survive, this alone can be used to prevent their spread using vertical resistance in a host-host-pathogen system (for obligate parasites). The vertical resistance needs to be in the second host which receives inoculum from the first host that lacks vertical resistance. The second host could be a new cultivar with new vertical resistance before it is widely grown and succumbs to the disease; the boom and bust of 'temporary' resistance such as *R1* for resistance to late blight of potatoes. The second host could be a spring cereal with vertical resistance grown north of a winter one where the pathogen spreads from south to north, such as the *Sr<sub>6</sub>* gene for resistance to stem rust (*Puccinia graminis*) in red spring wheat in North America, which was described in detail by Van der Plank. The resistance is durable as long as the gene is not used in the first host. Van der Plank thought that the best use of *R*-genes was in a multiline-pathogen system because the genes could not be mismanaged. He also pointed out that vertical resistance can work in a saprophytic medium-host-pathogen system where the host is an annual and the pathogen has a saprophytic life as well as a parasitic life; provided increased virulence reduces the ability of the pathogen to survive saprophytically. One of Van der Plank's examples of durable resistance was major gene resistance to cabbage Fusarium wilt, a worldwide disease of cabbage (*Brassica oleracea*), caused by the fungus *Fusarium oxysporum* f. sp. *conglutinans*. Although two races have been found to infect cabbage, inbuilt resistance is still considered the most effective method of control for this typical soil borne disease that is hard to manage with chemicals and crop rotation. One single dominant resistance gene *FOC1* has recently been mapped, molecular markers developed for molecular-marker-assisted breeding, and a putative TIR-NBS-LRR type candidate gene identified (Lv et al. 2014). It confers resistance to race 1 which is the only one found worldwide and for which major gene resistance has been effective since the 1930s; the second race is still confined to parts of the USA and Russia.

Van der Plank's theory also requires genes for vertical resistance to be quantified for strength. The complementary race that can overcome a strong gene experiences strong selection against it in the absence of the gene, whereas the complementary race to a weak gene experiences weak or even no selection. As an example, of the *R*-genes from *Solanum demissum* for resistance to late blight of potato, 1–3 were considered strong by Van der Plank, whereas 4–9 were considered weak and ineffective for controlling the disease (10 and 11 were discovered later). Interestingly, however, an extensive review of available data by Świeżynski et al. (2000) revealed that the resistance of differentials R5, R8 and R9 was overcome least frequently, and hence these *R*-genes may still be of value in some situations. In contrast, as horizontal resistance does not exert selection pressure on the pathogen to become more aggressive (the pathogen would do this anyway if possible), selection for improved horizontal resistance *per se* should be effective. Van der Plank (1968) did not regard genes for horizontal resistance as specialized resistance genes, but rather part of the general metabolism of plants.

## ***Horizontal Resistance***

Over 20 years after Van der Plank published *Disease Resistance in Plants*, Simmonds (1991) reviewed 200 publications on horizontal resistance. These provided examples from cereals, tubers, legumes, vegetables, fruits, ornamentals and miscellaneous crops of horizontal resistance to airborne fungal diseases and those caused by soil fungi, bacteria, viruses and animals; as well as examples of multiple resistances to different diseases. Simmonds (1991) thought that horizontal (quantitative) resistance was polygenically controlled and must be studied by biometrical-genetic methods, that it was pathotype (race)-non-specific and essentially durable, and that it had four components for reducing the rate of disease progress, as recognized by Van der Plank (1968). These are reducing the proportion of spores (of all races) that initiate lesions, increasing the latent period, reducing the abundance of spore production and reducing the amount (removal) of infectious tissue. Simmonds came to the very positive conclusions that horizontal resistance is universally available, usually highly heritable and responsive to selection, already keeps numerous crop diseases down to acceptably low levels and has socio-economically attractive features that are likely to increase its use in the future. In particular, it offers long-term stability of performance that must be valuable to small farmers in developing countries and is environmentally attractive because successful breeding programmes minimize the need for environmentally damaging chemical protectants. Not surprisingly then he thought that breeders should make greater use of the horizontal resistance universally available in cultivated species, the corollary being that vertical resistance is only of use in special circumstances and hence its introgression from wild relatives may not be required. Others, however, still find vertical resistance an attractive proposition.

## **Screening for Resistance**

Reliable testing is required both for screening germplasm collections for new sources of resistance and for assessing material under selection in a breeding programme. A reliable test for qualitative resistance is one where there are no escapes; susceptible control plants should all be susceptible. A reliable test for quantitative resistance is one in which there is little environmental variation or errors of measurement. A statistically significant offspring-midparent regression, for example, is evidence of heritable resistance, and the slope of the regression line is a measure of (narrow-sense) heritability. When low heritabilities result from environmental and genotype  $\times$  environmental interaction variation, heritability and response to selection can be increased through increased replication in individual tests and an increased number of tests, respectively, if this is thought to be a good use of additional resources. For any specific disease test, the best stage of plant growth for inoculation and scoring, inoculation concentration, method of

inoculation, and environmental conditions (temperature, day-length, humidity) are all factors to be determined by experiment, if not already known in the literature. Testing will need to use the most appropriate isolate or isolates of the pathogen (or population of pest), taking into account its differentiation into species, subspecies, biovars or physiological races; and the distribution of these in the target environments of the breeding programme. For some pathogens a mixture of isolates/inoculum from different geographical locations may be appropriate, for others the most virulent or most aggressive isolate may be required.

Different resistance mechanisms may occur in different parts of the plant, thus necessitating separate disease tests. Resistance to late blight of potato in the foliage does not guarantee resistance in the tubers, and vice versa, although correlations can be found in some germplasm (Stewart et al. 1994). Furthermore, for diseases such as late blight of potato, it is well known that resistance is often associated with late maturity and hence it is important to score both if one wants to break the association, and probably best to test early and late maturing germplasm separately (Bradshaw et al. 2004). With pest resistance, tolerance may be required in addition to resistance to ensure maximization of yield as well as control of the pest. Host genotypes that are resistant to potato cyst nematode attack are those that inhibit or reduce nematode multiplication rates. The roots of such genotypes are nevertheless invaded by juvenile nematodes and can suffer damage, with large yield losses being sustained when they are grown in infested soil. Genotypes differ in the degree of yield loss sustained in comparable situations, and these differences are referred to as differences in tolerance (Dale et al. 1988). Higher priority has, however, been given to resistance breeding and work on tolerance limited to identifying methods of screening potential new cultivars in their final stages of evaluation. The literature on screening for resistance is vast and a breeder will need to consult it for those diseases and pests of particular interest in his or her programme. Bradshaw et al. (2000) gave three contrasting examples from the potato programme at the former Scottish Crop Research Institute which will be repeated here by way of illustration. They emphasize the amount of work required to develop and use reliable disease tests.

### **Late Blight in the Foliage of Potatoes**

Foliage blight is typical of an air-borne fungal disease, although strictly speaking it is caused by an oomycete (*Phytophthora infestans*). The site used is in Ayrshire on the western seaboard of Scotland where the moist weather favours the development of late blight. The trial is planted in early May. Every third drill is the susceptible cultivar King Edward so that every test drill is adjacent to one spreader drill of King Edward. A randomized complete block design is used for each set of test material, with two or three replicates of two or three plant plots. Infector plants of King Edward are raised in pots in a glasshouse and inoculated in mid July with an appropriate complex race of *P. infestans*, assuming assessment is for quantitative



**Fig. 18.2** Assessment of resistance to late blight in the foliage of potatoes in a field trial in Ayrshire, Scotland (every third drill is the susceptible cultivar King Edward). (a) Genetic experiment in August 1998. (b) Breeding material in 2008

resistance. Three to four days later these infector plants are placed at metre intervals along the spreader drills. The amount of blight present on each test plot is scored on a 1–9 scale of increasing resistance (Cruickshank et al. 1982) at intervals of a few days during August (Fig. 18.2). The best scoring date is chosen retrospectively as either the one when the standards were showing their expected reaction to the disease on past experience of the trial, or the one that best discriminated between the material under test. In genetic experiments it is also possible to estimate the area

under the disease progress curve and the infection rate, as we shall see in the next section. The test is quite severe as the spreader rows can produce a lot of inoculum and interplot interference is a potential problem, as a clone with good quantitative resistance can by chance have very susceptible ones as neighbours. Nevertheless, the test has proved its worth over many years.

### **Powdery Scab of Potatoes**

Powdery scab is typical of a disease caused by a soil-borne fungus (*Spongospora subterranea*) that is difficult to work with because of variable inoculum levels in a field, coupled with variable viability of spore balls and specific infection conditions. Hence tests are done in special beds (Wastie et al. 1988a) or nurseries where dried peelings from infected tubers in the previous season are spread uniformly, and irrigation provided at critical periods. The number of infected tubers in each plot (i.e. tubers with at least one recognizable pustule or canker) is recorded at the end of the growing season, and the percentage of infected tubers calculated. In New Zealand, the breeder uses a disease index that takes severity of infection into account by counting the number of tubers in each of four disease categories (Russell Genet, personal communication).

### **Gangrene of Potatoes**

Gangrene caused by the fungus *Phoma foveata* is typical of a tuber-borne disease that develops in store. Appropriate-sized plots of the test material are raised free of disease in replicated trials in a field on an experimental farm and tubers harvested at maturity in early September. Tubers are stored in boxes, one for each plot, at 6–10 °C until mid November each year, when they are washed and inoculated with a mixture of isolates of *Phoma foveata* by a cornmeal-sand method (Wastie et al. 1988b). The tubers are dampened by being lightly watered, and are then rolled in a box of the diluted inoculum so that they are completely covered in sand, some of which adheres to them when they are removed. They are then replaced in the boxes and immediately incubated at 4 °C and 100 % RH for 8–10 weeks after which the level of infection is scored. In order to register a wide difference in score between resistant and susceptible genotypes, the number of tubers infected to a predetermined amount is recorded. An amount of either one-16th, one-eighth, or one-quarter of the surface area of the tuber covered with gangrene lesions is chosen according to the rapidity with which the infection has developed in the control cultivars at the time of scoring. The percentage of infected tubers in each sample is then calculated (Stewart and Wastie 1989).

## Measurement of Resistance: Disease Progress Curves

We are now going to examine some simple theory for modelling disease progress curves, as found in the book by Van der Plank (1968). Late blight of potatoes features extensively in his book and we will be examining some more recent data. It is important to consider whether or not a model is appropriate for a particular disease. We will assume that an epidemic is started by a low concentration of spores so that they act independently and the number of lesions produced is proportional to the number of spores (Van der Plank provides evidence for this). For example, inoculum of *Phytophthora infestans* (cause of late blight) can survive winter in cull piles of diseased tubers and then start an epidemic in the following crop. In our simple theory we are going to relate the infection rate of an epidemic to the proportion of infected tissue. Although this will prove adequate for our purpose, for late blight we should really be estimating infection rate based on infectious tissue. As a lesion of potato blight grows through a leaf following successful infection by a spore, there is an outer zone of mycelium working its way into healthy tissue. Behind it is a zone in which sporangia are formed if the humidity is high enough. In turn, behind this sporulating zone is a sterile, inner zone, containing perhaps some undispersed sporangia but producing no new sporangia even under optimal conditions of temperature and humidity. Thus there is a latent period before newly infected tissue becomes infectious. Then follows an infectious period before the tissue is removed from the epidemic because it neither produces inoculum nor is it susceptible to re-infection by *P. infestans*. Considerable necrosis in the centre of a lesion and a narrower zone of spores is a known characteristic of many resistant potato cultivars. Hence Van der Plank (1968) thought that resistance in the form of a reduced infectious period, as well as increased latent period, was probably important for late blight of potatoes, but of minor importance in the cereal rusts where the infectious period is 10 days or more. Nevertheless, our simple model will prove adequate for discussing both late blight and rusts.

### Theory

The infection rate ( $r$ ) of an epidemic is defined by the following differential equation:

$$\frac{dy}{dt} = ry(1 - y),$$

where  $y$  is the proportion of disease (e.g. leaf area of plants covered by lesions such as those of late blight of potatoes) at time  $t$  and  $(1-y)$  is the proportion of tissue still healthy and available for infection. (Van der Plank used  $x$  instead of  $y$ , but the reader will probably be used to  $y$  as the vertical axis in a graph). The differential equation is a mathematical model of an epidemic in which  $y$  increases with  $t$ , and

the rate of increase depends on both the amount of disease present and the amount of healthy tissue remaining to be infected. We solve the differential equation as follows.

$$\int dy/[y(1-y)] = \int dy/y + \int dy/(1-y) = r \int dt$$

Therefore

$$\ln(y) - \ln(1-y) = rt + k \text{ where } k \text{ is a constant, and}$$

$$\begin{aligned} \ln(y/(1-y)) &= rt + k, \\ y/(1-y) &= e^{rt+k} \\ y &= e^{rt+k}/(1 + e^{rt+k}) \end{aligned}$$

Hence the graph of  $\ln(y/(1-y))$  against  $t$  is a straight line of slope  $r$  and intercept  $k$ , whereas the graph of  $y$  against  $t$  is a curve, the disease progress curve.

Early in an epidemic,  $y$  will be small and  $(1-y)$  will be very close to 1, so that  $\ln(y) \approx rt + k$ , and

$\ln(y_1) - \ln(y_0) = rt_1 - rt_0$ , where  $y_0$  and  $y_1$  are the proportions of disease at times  $t_0$  and  $t_1$ .

Hence

$\ln(y_1/y_0) = r(t_1 - t_0)$  or  $(t_1 - t_0) = (1/r)\ln(y_1/y_0)$ , so that  $(t_1 - t_0)$  is the time required for the amount of disease to increase from  $y_0$  to  $y_1$ .

Also

$$y_1/y_0 = e^{r(t_1 - t_0)}, \text{ so that}$$

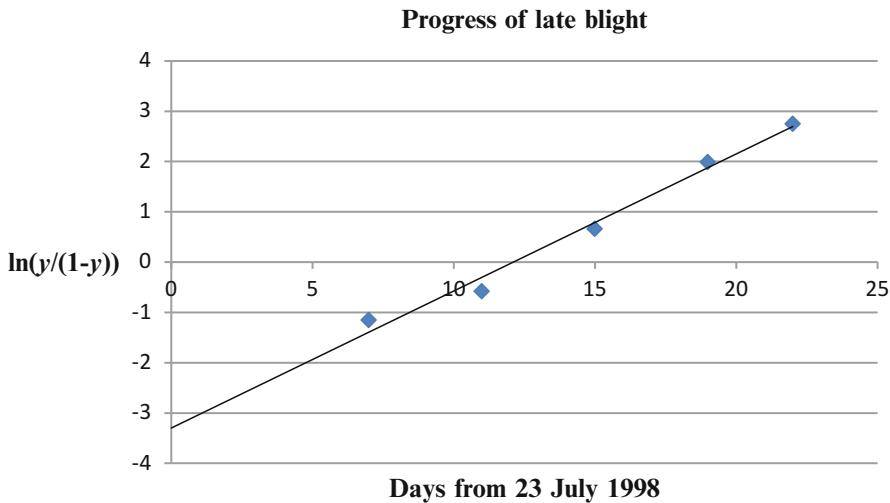
$$y_1 = y_0 e^{r(t_1 - t_0)}$$

In other words, early in the epidemic we have exponential growth in which infections are independent of one another (i.e. no interference), as observed with diseases such as late blight of potatoes and the cereal rusts. Furthermore, these equations show that the longer an epidemic lasts (the bigger  $t - t_0$ ), the greater the effect of a given value of  $r$  (horizontal resistance) on the final value of  $y$  for a given initial value  $y_0$ .

It is now time to use these results to construct and examine some disease progress curves.

## Practice

Bradshaw et al. (2004) made field assessments of foliage resistance to late blight of potatoes in Ayrshire in 1998, as described in the previous section. On July 20, glasshouse-grown plants of cultivar King Edward were inoculated in the laboratory



**Fig. 18.3** Progress of complex race 1, 2, 3, 4, 6, 7 of *Phytophthora infestans* on six R-gene differentials (mean of R1, R2, R3, R4, R6 and R7) of potato in Ayrshire, Scotland in 1998, where  $y$  is proportion of disease

with a zoospore suspension of the complex race 1, 2, 3, 4, 6, 7 of *P. infestans*. Three days later, on July 23, they were placed in the spreader rows of the trial. The amount of blight present on the two (or three) plants in each plot was scored on a 1–9 scale of increasing resistance on July 30 and on August 3, 7, 11 and 14. The 1–9 scale corresponded to the percentage of necrotic tissue (100, 90, 80, 70, 60, 40, 25, 10 and 0). Two replicates of the R-gene differentials, with the exception of *R*<sub>9</sub>, which was not available, were planted adjacent to the trial in three-plant plots and scored on the same dates. The scores confirmed that the trial had been inoculated with the intended race. Despite some differences, I am going to assume that the R1, R2, R3, R4, R6 and R7 differentials were equally susceptible to race 1, 2, 3, 4, 6, 7 of *P. infestans* and work with their mean scores. Figure 18.3 shows the plot of  $\ln(y/(1-y))$  against  $t$  in days from July 23. It is a straight line of slope ( $r$ ) 0.2725 with intercept ( $k$ ) -3.299. In other words:

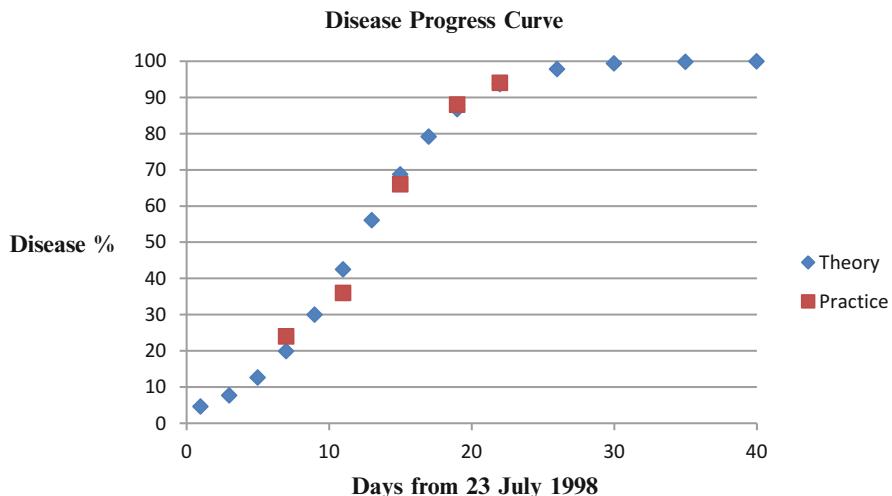
$$\ln(y/(1-y)) = 0.2725t - 3.299, \text{ where}$$

$r = 0.2725$  is the rate of infection.

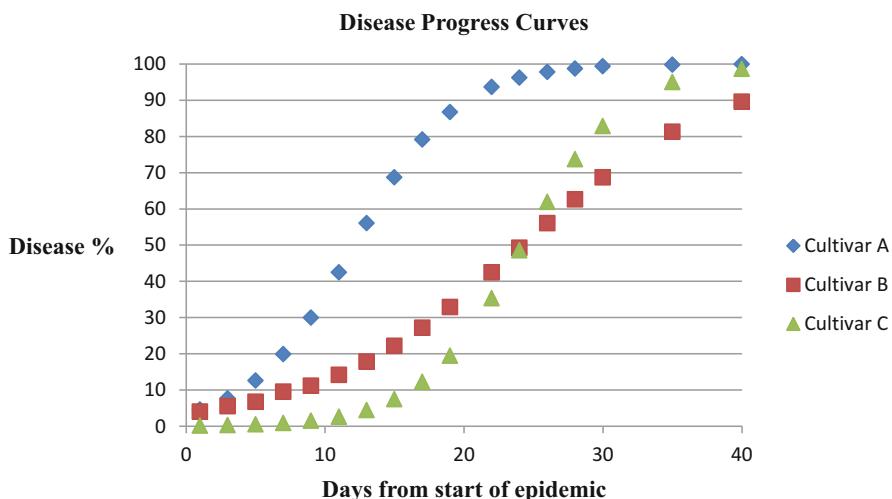
These values for  $r$  and  $k$  can be used to plot the disease progress curve shown in Fig. 18.4 from:

$$y = e^{rt+k} / (1 + e^{rt+k}),$$

where  $y$  is multiplied by 100 to get the percentage of disease at time  $t$  in days.



**Fig. 18.4** Disease progress curve with  $r=0.2725$  and  $k=-3.299$  and five actual values for complex race 1, 2, 3, 4, 6, 7 of *Phytophthora infestans* on six R-gene differentials (mean of R1, R2, R3, R4, R6 and R7) of potato in Ayrshire, Scotland in 1998



**Fig. 18.5** Theoretical disease progress curves differing in  $r$  and  $k$  (cultivar A:  $r=0.2725$  and  $k=-3.299$ ; cultivar B:  $r=0.13625$  and  $k=-3.299$ ; cultivar C:  $r=0.2725$  and  $k=-6.598$ )

Despite only having five actual values, the disease progress curve is typical of late blight and shows why it is potentially one of the most destructive diseases in the world. Under favourable conditions, it took about 3 weeks to destroy the susceptible plants. We can now explore two questions about disease progress curves, the effect of changing  $r$  and the effect of changing  $k$ . In Fig. 18.5, the rate of infection on

cultivar B is half that on cultivar A ( $r=0.13625$  compared with  $r=0.2725$ ;  $k=-3.299$  for both) and  $k$  for cultivar C is double that of cultivar A ( $k_C=-6.598$  compared with  $k_A=-3.299$ ;  $r=0.2725$  for both). Halving the rate of infection halves the rate of progress of the disease, whereas the change in  $k$  ( $k_A-k_C$ ) has delayed the epidemic by 12 days [ $(k_A-k_C)/r=12.1$ ]. So what does the change in  $k$  represent? The answer is the amount of disease at the start of the epidemic when  $t=0$  and  $y$  is very small. Then we have  $\ln(y_A)=k_A$  and  $\ln(y_C)=k_C$  so that  $(k_A-k_C)=\ln(y_A)-\ln(y_C)=\ln(y_A/y_C)$ . But we have shown above that with the same rate of infection  $r$ ,  $\ln(y_A/y_C)/r$  is the time required for the level of disease to increase from  $y_C$  to  $y_A$ . Furthermore, if we assume that the number of lesions produced (amount of disease) is proportional to the number of spores starting the epidemic, we conclude that the epidemic is delayed by fewer spores infecting cultivar C than cultivar A. The ratio is in fact 27.1 to 1. This situation could arise if a natural epidemic was started by a mixture of races (1, 2, 3, 4, 6, 7) and (1, 2, 3, 4, 6, 7, 10) in the ratio of 27 to 1 and cultivar A possessed *R*-genes *R1R2R3* and cultivar C possessed *R10*. All spores could infect cultivar A but only 1 in 28 could infect cultivar C. The number of spores infecting a cultivar is determined by the race composition of the pathogen and whether or not the cultivar possesses resistance to any of the races.

In summary, the amount of disease is determined by the number of spores infecting a cultivar and the subsequent rate of infection. Vertical resistance affects the former and horizontal resistance the latter. Hence vertical resistance delays the start of an epidemic whereas horizontal resistance slows it down after it has started. Van der Plank (1968) gave examples for late blight of potatoes from the 1950s. He presented data from West Germany in 1955 showing that the *R1* gene delayed epidemics by 5 days and the *R3R4* combination delayed them by 15 days. He also presented data from 117 potato fields in the Netherlands that compared the rate of disease progress on susceptible cultivar Bintje ( $r=0.42$  per day), intermediate cultivar Eigenheimer ( $r=0.21$  per day) and resistant cultivar Voran ( $r=0.16$  per day). The two types of resistance can be combined, but how easy are they to recognize in practice? We need to distinguish disease nurseries that rely on natural epidemics, those that have been inoculated with a mixture of races, and those inoculated with a single race.

## Disease Nurseries

Let us take another look at cultivars A, B and C in Fig. 18.5, where A is susceptible, B has horizontal resistance, C has vertical resistance and an appropriate mixture of races has been used. It is common practice for breeders to assess resistance when the susceptible control(s) is clearly susceptible and there is good discrimination between resistant and susceptible germplasm. On day 19 (fourth score on August

11) the diseases percentages are 87, 33 and 19 %. Cultivar A is clearly susceptible and cultivars B and C resistant, with C apparently more resistant than B. But if the assessment is delayed to day 24, cultivars B and C now both have 49 % disease compared with 96 % for cultivar A and appear equally more resistant (or less susceptible) than the susceptible control. A common way to try to make use of the information from a number of scoring dates is through the area under the disease progress curve (AUDPC). Again the areas from day 0 to day 19 or day 24 would show the order A > B > C, but extending the areas beyond day 24 would reduce the difference between B and C. The exact areas can be calculated from the definite integral of the disease progress curve from time zero to  $t$ , but a good approximation with only a few scores such as the five in Fig. 18.4 is as follows (Shaner and Finney 1977).

$$\text{AUDPC} = \sum [(y_{i+1} + y_i)/2] [t_{i+1} - t_i] \text{ from } i = 1 \text{ to } n,$$

where  $y$  is the disease percentage (or other unit of severity) at the  $i$ th observation,  $t$  is the time (in days) at the  $i$ th observation and  $n$  is the total number of observations. The first observation in Fig. 18.5 would be  $y=0$  at  $t=0$ . Sometimes areas are expressed relative to the maximum possible area.

Neither the best date for discriminating between resistant and susceptible genotypes nor the area under the disease progress curve can discriminate between horizontal and vertical resistance in the way achieved by the graph of  $\ln(y/(1-y))$  against  $t$ . Even here there are potential complications when comparing genotypes. Genotypes may combine both types of resistance, changes in environmental conditions and the physiological state of the plants may occur during an epidemic so that the rate of infection changes with time, and a new race may arrive during the course of an epidemic. So can a high level of horizontal resistance be distinguished from vertical resistance, and selected if desired? Andrivon et al. (2006) have provided evidence from natural late blight infections in potato breeding nurseries in France, that the graph of  $\ln(y/(1-y))$  against  $t$  can be used to evaluate potential cultivars for their ability to delay the onset of an epidemic and also their ability to reduce infection rates.

White and Shaw (2009) assessed potential cultivars of potato at a site conducive to late blight on the Isle of Anglesey, North Wales from 2005 to 2008. During this period, the population of *P. infestans* in Great Britain changed. Instead of recording a high frequency of A1 mating type isolates, surveys detected increasing frequencies of the A2 mating type due primarily to a single A2 strain with resistance to metalaxyl, known as Blue-13 (SSR genotype 13\_A2), which was first detected in 2005 and was the dominant strain in 2008. It increased in frequency in Britain from 12 % in 2005 to 71 % in 2007 (Cooke et al. 2008) and is particularly virulent and aggressive. The trials in 2005 and 2006 relied on spontaneous infection with *P. infestans* (not strain Blue-13) whereas in 2007 and 2008 a single plant of susceptible cultivar Bintje was planted in each plot and inoculated with an isolate of strain Blue-13. In all four trials, foliage of the

susceptible standard Bintje was destroyed by blight within 10 days of the appearance of the first symptoms. In 2005 (and 2006) the resistant standards, cultivars Lady Balfour and Robijn, showed a delay of disease initiation (compared with Bintje) of more than 20 days until the disease progressed as rapidly as happened on Bintje. In contrast, cultivars Sárpo Mira and Axona remained virtually free of disease. The relative areas under the disease progress curves were 0.81 (Bintje), 0.29 (Lady Balfour), 0.30 (Robijn) and 0.00 (Sárpo Mira and Axona). In 2008 (and 2007), the delay in disease initiation on both Lady Balfour and Robijn was only 7 days, followed by a rate of increase almost as high as on Bintje. The Sárpo cultivars also showed a shorter delay but their subsequent disease progression was much less than that of the other cultivars. Fifty six days into the epidemic, the foliage of Sárpo Mira was only 40 % blighted. The relative areas under the disease progress curves were 0.88 (Bintje), 0.81 (Lady Balfour), 0.75 (Robijn), 0.53 Axona and 0.25 (Sárpo Mira). The resistance of Lady Balfour and Robijn appeared to be mainly vertical whereas that of Sárpo Mira and Axona was largely horizontal. The relative areas under the disease progress curves in 2005 and 2006 appeared to largely reflect vertical resistance but in 2007 and 2008 horizontal resistance. Nevertheless, the authors presented results across all four trials as relative areas under the disease progress curves.

Where a breeding programme aims to select for quantitative resistance, it is common practice to inoculate a disease nursery with a complex race capable of overcoming any *R*-genes known to be present, as described in the previous section for late blight of potatoes. Mixtures of races will give biased results unless each can overcome all of the *R*-genes present, as pointed out by Van der Plank (1968). Assessments of resistance are then either made on the best date for discriminating between resistant and susceptible genotypes or as the relative areas under the disease progress curves; both of which should largely reflect the rate of increase of disease. However, Simmonds and Wastie (1987) argued in favour of the mean resistance over all scoring dates as a simple and accurate way of making use of all of the data. The alternative way of selecting for quantitative resistance is to create *R*-gene free germplasm so that screening can be done with an isolate of any race. This approach was found useful by CIP in developing their potato Populations B1, B2 and B3, mentioned under base broadening in Chap. 15. Race-specific resistance to late blight was eliminated from these populations where necessary by progeny testing with crosses to a susceptible tester that did not possess *S. demissum* derived *R*-genes for resistance (Trognitz et al. 2001).

## Examples of Relevance to Breeding for Durable Resistance

It is now time to provide a series of examples of relevance to breeding for durable resistance before summarizing the issues that have emerged. We will consider recent research on resistance to late blight of potatoes, slow-rusting wheats, Mlo resistance to powdery mildew in barley, the possibility of broad spectrum bacterial

resistance and the possibility of selecting for durability of a virus resistance gene. The set of examples are not comprehensive in terms of crops and diseases, but have been chosen to stimulate thought about strategies and issues.

## A Further Look at Resistance to Late Blight of Potatoes

Lees et al. (2012) confirmed that the advent of *P. infestans* genotype 13\_A2 (Blue-13) in the UK resulted in previously resistant cultivars such as Lady Balfour and Stirling (they share same late blight resistant parent) now being classified as susceptible (resistance ratings dropped from 8 to 4, on a 1–9 scale of increasing resistance), whereas Sárpo Mira was still resistant (rating 8). They did field trials in Ayrshire as already described. Interestingly Roslin Eburu, released in 1960 for use in Kenya (Bradshaw 2009a), was still resistant (rating 7–8) and the new cultivars Bionica and Toluca (both with resistance gene *Rpi-blb2* introgressed from *S. bulbocastanum*) were also resistant (rating 7). Roslin Eburu has one of the same *S. demissum* derived sources of resistance as Pentland Dell, which succumbed to blight in 1967. Lady Balfour has been grown in the UK as an organic cultivar since 2001, and is still a successful organic cultivar despite now being susceptible to late blight. Sárpo Mira is grown by gardeners in the UK for its late blight resistance, but Lady Balfour, Stirling and Sárpo Mira have never been grown extensively on a commercial scale in the UK. Where as in the UK susceptible cultivars are widely grown under the protection of fungicides, the deployment of resistance genes is not the driver for selection of new genotypes of *P. infestans*. Indeed, it almost seems a matter of chance whether or not a cultivar is resistant to a new genotype of *P. infestans*. Nevertheless, it is worth looking at the genetic analyses of the resistance of Stirling and Sárpo Mira as both cultivars were selected for field resistance, the former in Scotland and the latter in Hungary. The results suggest that the relationships between qualitative resistance and quantitative field resistance and their durability are not as clear cut as once thought, and that breeders may need to think in terms of assembling the best combinations of *R*-genes and major QTLs for resistance, in the light of predictions about the likely responses of populations of *P. infestans* to deployment of these combinations. Neither the analysis of Stirling nor the analysis of Sárpo Mira provides evidence of field resistance that does not display resistance × isolate interactions, that is horizontal resistance as proposed by Van der Plank (1968). It remains to be seen if analysis of other sources of field resistance, such as CIP's Populations B1 to B3 dealt with in Chap. 15, can provide such evidence. High yielding clones with good field resistance have been selected from population B3C2 in Uganda, after two cultivars from population B3C1 succumbed to late blight (Byarugaba et al. 2013).

## ***Resistance of Stirling***

The late blight resistance of cultivar Stirling was analyzed before the advent of *P. infestans* genotype 13\_A2 (Blue-13) by assessing the resistance of offspring from the cross with the susceptible clone 12601ab1 (Stewart et al 2003; Bradshaw et al. 2004). The provisional map positions and effects of resistance genes were confirmed by further analyses of the results (Hackett et al. 2014). The results can be summarized as follows.

Whole-plant glasshouse tests revealed that Stirling has a single copy of a major *R*-gene on linkage group XI (chromosome 11) between 81 and 87 cM (distal end) which provided foliage resistance to simple race 1, 4 (restricted lesions or spreading lesions on the oldest leaves alone) but was overcome by complex race 1, 2, 3, 4, 6, 7 (at least a few spreading lesions on leaves of any age). The gene is close to or part of a large cluster of NB-LRR genes including *R3a*. Stirling does have the *R7* differential in its pedigree but the *R*-gene has not been confirmed as *R7*. Further results from the cross of Stirling with susceptible cultivar Maris Piper confirmed that the *R*-gene provides resistance to the simple race but not the complex race (Soloman-Blackburn et al. 2007).

The defeated *R*-gene, or a closely linked gene(s), provided a small but statistically significant contribution to the foliage field resistance of Stirling, as assessed in Ayrshire in 1998 (average of five scoring dates). The 137 clones with the *R*-gene had an average score of 6.27 compared with 5.78 (difference  $0.49 \pm 0.170$ ) for the 78 clones without the *R*-gene; this difference being 9.4 % of the range found in the progeny clones (2.70–7.90, on a 1–9 scale of increasing resistance). Results from the cross of Stirling with susceptible cultivar Maris Piper and the backcross of clone 94B13A57 to Maris Piper provided further evidence of the defeated *R*-gene contributing to resistance, this time in a glasshouse test (Soloman-Blackburn et al. 2007). The average scores of the backcross clones with and without the *R*-gene were 4.23 and 2.52.

Stirling has a single copy of a gene for early maturity on linkage group V (chromosome 5) between 14 and 22 cM that is associated with susceptibility to blight. Clones with the gene had an average score of 3.77 compared with 5.63 for those without the gene on the fourth scoring date in Ayrshire in 1998. They were also more susceptible to blight in their tubers (74.63 versus 45.17 % susceptible) in a glasshouse test in 1999.

Stirling possesses two copies (duplex) of a resistance allele (*R*) on linkage group IV (chromosome 4) between 22 and 30 cM. Offspring with genotype *RRrr* had an average fourth score of 6.3, those with *Rrrr* 5.1, and those with *rrrr* 1.7. The corresponding (glasshouse) means for tuber blight were 38.2, 57.1 and 84.1 %. Maturity did not affect these results for chromosome 4 which confirmed that the gene was a true resistance gene. Segregation at this locus explained 43.6 % of the variation in foliage blight scores and 19.2 % of the variation in tuber blight scores.

A clone (94B13A61) from the cross of Stirling with Maris Piper inherited field resistance but not the major *R*-gene from Stirling. It is reasonable to assume that the field resistance was due to a resistant allele from chromosome 4 (Soloman-Blackburn et al. 2007). When 11 offspring selected for resistance from the cross between 94B13A61 and Maris Piper were assessed in Ayrshire in 2007, they were on average as susceptible as Maris Piper and none was more resistant than Stirling. That year *P. infestans* genotype 13\_A2 (Blue-13) was naturally present, dominated the blight trial, and proved virulent (compatible) on the allele from chromosome 4. We thus appear to have a QTL × isolate interaction (Bradshaw 2009c).

### ***Resistance of Sárpo Mira***

Tomczynska et al. (2014) examined 137 offspring from the cross between Sárpo Mira and susceptible cultivar Maris Piper in detached leaflet assays using complex races of *P. infestans*. They mapped a marker (45/XI) at the end of Sárpo Mira's chromosome XI, in the vicinity of the *R3* locus, which explained between 55.8 and 67.9 % of the variation in the mean scores for late blight resistance. They concluded that they had probably found a new *R*-gene homologue, located in the *R3* gene cluster, that may correspond to gene *Rpi-Smira1* identified by Rietman et al. (2012). Orłowska et al. (2012) reported that in Denmark, when Sárpo Mira was crossed with susceptible cultivars, a 1:1 segregation of resistant to susceptible offspring occurred, as expected from the segregation of a major *R*-gene. However, the work of Rietman et al. (2012) identified at least five resistance genes in Sárpo Mira, four of which (*R3a*, *R3b*, *R4* and *Rpi-Smira1*) conferred qualitative resistance that could be detected in detached-leaf tests, and a fifth, *Rpi-Smira2*, which was only detected under field conditions. They used co-segregation of hypersensitive response data obtained with use of RXLR effectors (effector assays) and phenotypic resistance data obtained with differential *P. infestans* isolates to reach these conclusions. Functionally testing effectors of *P. infestans* for inducing cell death responses in Sarpo Mira proved a powerful technique. Races of *P. infestans* are common that can overcome *R3a*, *R3b*, and *R4* and this explains why a 1:1 segregation can be found when using the appropriate complex races. Races do also exist that can overcome *Rpi-Smira1*. The authors suggested that field resistance conferred by *Rpi-Smira2* may represent a weaker form of *R*-gene based resistance that does not cause a full-blown hypersensitive resistance and resistance phenotype upon recognition of the pathogen. There is still much to learn about the components of resistance of Sárpo Mira and their durability, and hence for how long and in what circumstances it will remain a useful source of resistance genes.

## Slow-Rusting Wheats

Singh and Trethowan (2007) provided an account of CIMMYT's (International Maize and Wheat Improvement Centre, Mexico) spring bread wheat breeding programme for irrigated and rainfed production systems of the developing world. It included breeding strategies to safeguard wheat crops from important diseases. They pointed out that breeding for resistance to rust and powdery mildew fungi was particularly challenging due to the evolution of new virulence through migration, mutation, or recombination of existing virulence genes. As a consequence, the emphasis at CIMMYT has been on selection for general (race-nonspecific) resistance which for the rusts is referred to as slow-rusting. Genetic analysis has revealed that adult plant resistance to brown leaf rust (*Puccinia triticina*) is based on the additive interaction of gene *Lr34* and two or three additional slow-rusting genes. Together they reduce disease severity to 1–5 % compared with 10–15 % for *Lr34* plus one or two additional minor genes, 40 % for *Lr34* alone and 100 % in a susceptible cultivar. The presence of *Lr34* is indicated by leaf tip necrosis in adult plants. A second slow-rusting resistance gene *Lr46* was identified more recently and at least 10–12 slow-rusting genes are involved in the adult-plant resistance of CIMMYT wheat germplasm. Likewise for yellow stripe rust (*Puccinia striiformis*), combinations of *Yr18* (completely linked to *Lr34*) and 3–4 additional slow-rusting genes result in adequate resistance levels in most environments; and recently identified *Yr29* is completely linked to *Lr46*. The occurrence of chlorotic or necrotic stripes can make distinguishing slow-rusting resistance from race-specific resistance difficult.

Black stem rust (*Puccinia graminis*) had been controlled effectively by genetic resistance in cultivars associated with the Green Revolution of the 1960s and 1970s. Although the slow-rusting gene *Sr2* conferred inadequate protection when present alone, in combination with other unknown slow-rusting genes the resulting complex provided the foundation for durable resistance to stem rust in germplasm from the University of Minnesota in the United States, Sydney University in Australia, and the spring wheat germplasm developed by Dr. Borlaug at CIMMYT in Mexico. Furthermore, race-specific resistance gene *Sr31* (on rye translocation 1B.1R) contributed to high levels of resistance in several cultivars developed worldwide. However, detection in 1998 in Uganda of race Ug99 with broad virulence, including virulence for *Sr31*, and its subsequent spread to Kenya and Ethiopia led to the Borlaug Global Rust Initiative in 2005. There was justified fear that it would spread, either wind-mediated or human-aided, in Africa, and into the Middle East and Asia. Wheat cultivars grown on over 90 % of the world's wheat area were considered susceptible to Ug99. Perhaps not surprisingly, on Sunday 20 April 2014, *The Independent* newspaper in the UK carried the headline "Wheat rust: The fungal disease that threatens to destroy the world crop". Hopefully this assessment is premature. Singh et al. (2011) provided a review of the spread of the disease and progress in breeding for resistance since 2005.

By 2011 seven races belonging to the Ug99 lineage were known and had spread to various wheat-growing countries in the eastern African highlands, as well as Zimbabwe, South Africa, Sudan, Yemen and Iran. Screening in Kenya and Ethiopia had identified a low frequency of resistant wheat cultivars and breeding materials. Identification and transfer of new sources of race-specific resistance from various wheat relatives was underway to enhance the diversity of resistance. Although over 50 different stem rust resistant genes were known, including multiple alleles at three loci, only *Sr22*, *Sr26*, *Sr35*, and *Sr50* were considered effective against all current important races and therefore had the potential to be utilized successfully. Hence there was a need for the identification and use of new sources. Newcomb et al. (2013), for example, screened 2509 spring wheat landraces from the USDA collection in eight field seasons at the Kenya Agricultural Research Institute, Njoro, where the Ug99 race group had become endemic, and identified 246 resistant ones including Iranian accession PI626573. Subsequent high-density mapping located a single dominant gene *SrWLR* on chromosome 2B (Zurn et al. 2014). As this is an active area of plant breeding research, literature searches by the reader will reveal more genes that have been identified and mapped. Evaluation in field trials in Kenya had also confirmed previous observations that the slow-rusting gene *Sr2* conferred inadequate protection when present alone, but was an important component of effective adult plant resistance. The CIMMYT wheat improvement group is therefore focusing their efforts on breeding for adult plant resistance while other groups around the world are utilizing race-specific resistance (Singh et al. 2011). New Ug99-resistant cultivars that yield more than current popular cultivars are being released and promoted, but major efforts are required to displace current Ug99 susceptible cultivars with ones that have diverse race-specific resistance or adequate adult plant (field) resistance to mitigate the Ug99 threat.

### **Breeding Slow-Rusting Wheats**

Singh and Trethowan (2007) presented CIMMYT's spring wheat programme as a successful example of breeding for resistance to leaf and stripe rusts based on minor genes. Briefly, in the early 1970s selection in the segregating generations of breeding programmes was practised for plants and lines that showed at most 20–30 % rust severity. As a result, minor genes for slow-rusting became widely distributed within CIMMYT's spring wheat germplasm. By the early 1990s near-immune lines were being produced which combined four or five minor genes for resistance to both leaf and stripe rusts with high yield. Genes were brought together in 3-way (3-parent) and 4-way (4-parent) crosses, and plants in large segregating populations were selected under artificially created rust epidemics. Races of pathogens were used that had virulences for the race-specific resistance genes present in the parents. The resulting highly resistant lines were, and are, being used as parents in crosses with susceptible but well-adapted cultivars liked by farmers and hence grown across large areas. A single backcross-selected bulk scheme proved

successful in which the adapted cultivars provided 75 % of the genes in the first segregating generation. Each backcross ((res × sus) × sus) comprised 400–500 seeds and selection was practised for resistance and other agronomic features under high rust pressure in this and subsequent generations. As slow-rusting genes are partially dominant, backcross (BC<sub>1</sub>) plants carrying most of the genes showed intermediate resistance and could be visually selected. Selected plants naturally self-pollinated and about 1600 plants per cross were grown as spaced plants in the next generation (F<sub>2</sub>), and 1000 plants in subsequent generations (F<sub>3</sub> to F<sub>5</sub>). Selection was for low to moderate disease severity in early generations and low severity at F<sub>4</sub> and F<sub>5</sub>. One spike from each selected plant contributed to the bulk of the cross until F<sub>4</sub>, then at F<sub>5</sub> plants were harvested individually. The F<sub>6</sub> lines were evaluated in small plots and those selected advanced to yield trials. Lines were identified which combined high levels of resistance to leaf rust or stem rust or both with 5–15 % higher yield potential than the adapted cultivar. The breeding programme was relatively straightforward to operate on a large scale.

## Mlo Powdery Mildew Resistance in Barley

During the 1980s spring barley (*Hordeum vulgare*) cultivars with Mlo resistance to powdery mildew (*Blumeria graminis* f.sp. *hordei*) became extensively cultivated in Western Europe, and by 1990 were grown on about 700,000 ha (30 % of crop). This was a new kind of resistance that proved remarkably durable. It is effective only in living epidermal cells and hence not against pathogens that kill their host cells with toxins, nor against fungi that enter through the stomata and penetrate mesophyll cells. Jorgensen (1992) summarized knowledge about Mlo resistance and its utilization in 1992 as follows. It was originally described in a powdery mildew resistant barley mutant in 1942 and has been mutagen-induced repeatedly since then (over 150 mutants of which *mlo1* to *mlo10* are the best known). About 1970 it was also recognized in barley landraces collected in Ethiopia in the 1930s. The spontaneously occurring allele in the Ethiopian barley line Grannenlose Zweizeilige was designated *mlo11*. The resistance is unique in that (1) it does not conform to the gene-for-gene system; (2) *mlo* genes originating from different mutational events map as non-complementing recessive alleles in a single locus located distally on the long arm of barley chromosome 4; (3) all alleles confer the same phenotype, though with small quantitative differences; (4) it is effective against all isolates of the pathogen; and (5) the resistance is caused by rapid formation of large cell wall appositions (papillae) that prevent penetration by the infection peg of the fungus. Although powdery mildew isolates with elevated Mlo aggressiveness have been produced on barley in the laboratory, they have not been found in nature. The exploitation of Mlo resistance was hampered by pleiotropic effects of the *mlo* genes, namely necrotic leaf spotting and reduced grain yield, but these were overcome by further breeding work. This was stimulated in the 1970s and early 1980s by the realization that easier to use race-specific resistance (for example, the

*Mla* genes) was not durable. Although most research was done on mutagen-derived *mlo* resistance genes, it was three Ethiopian sources of *Mlo* resistance (all probably with *mlo11*) that were first introduced into commercial European cultivars. Subsequently Freialdenhoven et al. (1996) identified two unlinked genes (*Ror1* and *Ror2*) that are both required for the function of *mlo*. Six susceptible M<sub>2</sub> individuals were identified and isolated from chemically mutagenized seed carrying the *mlo5* allele by inoculation with a fungal isolate (K1). Susceptibility in each of these individuals was due to monogenic, recessively inherited mutations in loci unlinked to *mlo*. The authors concluded that the *Mlo* wild-type allele functions as a negative regulator and the *Ror* genes act as positive regulators of the non-race-specific resistance response. The spontaneous formation of cell wall appositions in *mlo* plants is suppressed in *mlo/ror* genotypes.

Since 1996 much research has been done on the mechanism and the utilization of *Mlo* resistance, but two examples will suffice to give an indication of the nature of this work. Reinstädler et al. (2010) analyzed novel induced *mlo* mutant alleles and used site-directed mutagenesis in combination with transient gene expression to reveal functionally important domains in the heptahelical barley *Mlo* protein encoded by the *Mlo* locus. The mutant alleles revealed amino acid residues that are critical for *Mlo* function as a seven transmembrane-domain protein in the context of powdery mildew susceptibility. The protein interacts with calmodulin to negatively regulate plant defence and promote susceptibility to powdery mildew fungi in both monocot and dicot plants. In contrast, McGrann et al. (2014) have found and analyzed a trade off between *Mlo* resistance to powdery mildew and increased susceptibility to a newly emerged disease of barley in Europe, *Ramularia* leaf spot, caused by the fungus *Ramularia collo-cygni*. *Mlo* resistance has in fact been shown to increase susceptibility to most of the facultative pathogens tested in recent years, but the reason is not yet understood.

## Broad Spectrum Bacterial Resistance

Plants can recognize potential pathogens via two perception systems. We have already met effector-triggered immunity (ETI) as it corresponds to gene-for-gene, race-specific resistance in which virulence effectors are recognized by intracellular resistance proteins encoded by *R*-genes. The other system detects conserved microbial molecules named pathogen- or microbe-associated molecular patterns (PAMPs or MAMPs), through pattern-recognition receptors (PRRs) leading to PAMP-triggered immunity (PTI). Although the overall importance of PAMP-triggered immunity for plant defence has been established, it has not been used to confer disease resistance in crops. Nevertheless, work done by Lacombe et al. (2010) in controlled laboratory conditions suggests that expression of PAMP recognition systems could be used to engineer broad-spectrum disease resistance to important bacterial pathogens, potentially enabling more durable and sustainable resistance in the field. They demonstrated that (transgenic) expression of EFR (a leucine-rich repeat

receptor kinase), a PRR from the cruciferous plant *Arabidopsis thaliana*, confers responsiveness to bacterial elongation factor Tu in the solanaceous plants *Nicotiana benthamiana* and tomato (*Solanum lycopersicum*), making them more resistant to a range of phytopathogenic bacteria from different genera as diverse as *Pseudomonas*, *Agrobacterium*, *Xanthomonas* and *Ralstonia*. PAMP-triggered immunity may be an important aspect of non-host resistance.

## Durability of a Virus Resistance Gene

Quenouille et al. (2014) studied the frequency of breakdown of resistance to *Potato virus Y* (PVY) in pepper (*Capsicum annuum*) carrying major resistance allele *pvr2<sup>3</sup>*. They created a segregating doubled-haploid population comprising 350 lines from a cross between an inbred line carrying allele *pvr2<sup>3</sup>* in a partially resistant background and an inbred line carrying susceptibility allele *pvr2<sup>+</sup>*. They came up with the interesting conclusion that combining a major resistance gene with quantitative resistance factors (QTLs) preserved the durability of the resistance gene. Resistance breaking frequency was determined by mechanically inoculating 30 seedlings of *pvr2<sup>3</sup>* lines with a 'CI chimera' of PVY that is not infectious *per se*, and 38 days latter evaluating the number of seedlings with systemic infection as a result of the presence of resistance breaking mutations. Quantitative resistance (to virus accumulation) due to genetic background was determined using a mutant of the CI chimera that could overcome the *pvr2<sup>3</sup>* allele and hence cause infection. The resistance breakdown frequency was a highly heritable trait and four QTLs explained 70 % of the variation in *pvr2<sup>3</sup>* breakdown frequency. Three of the four QTLs were also involved in quantitative resistance, strongly suggesting that QTLs controlling quantitative resistance have a pleiotropic effect on the durability of the major resistance gene. These intriguing results offer a new strategy in breeding for durable resistance that requires further exploration.

## Durable Resistance

A review of durable resistance by Mundt (2014) is as much a reminder of questions still to be answered as it is of progress in recent years. Advances in understanding the molecular mechanisms of host-pathogen interactions have demonstrated that avirulence genes are also effectors of pathogen virulence, with substantial redundancy among effector genes. But how can this knowledge help us to understand and attain durability of resistance? What is the mechanistic basis of quantitative resistance and is it relevant to its durability? In other words, do minor genes function through morphological and developmental processes and on basal host defence, or are they simply weak major genes that are likely to be overcome by loss-of-function mutations in the pathogen? And is minor gene resistance available against all

pathogens? A relatively small number of genes can account for a large proportion of quantitative resistance; but how many genes are sufficient, are components of resistance pleiotropically controlled, and will quantitative resistance select for increased pathogen adaptation and aggressiveness. Possible deployment strategies for major genes are well-known, but there are many unanswered questions, and choosing the best one is still difficult, assuming there is a choice. Deployment of a single major gene can still make sense in environments where a disease occurs at moderate or low severity, or where there is no other option. Rotation of resistance genes over time is difficult to manage in practice and virulences may not decline to their original frequencies once the corresponding resistance gene is removed. There is broad agreement that combining genes for resistance (pyramids) is a useful way to increase their durability, with many known successes such as the control of stem and leaf rusts of wheat. Resistance gene combinations have kept wheat stem rust in check since the mid-1950s, but the discovery of race Ug99 in Uganda in the late 1990s now threatens these pyramids. Elucidating the mechanisms by which pyramids provide durability is important because it is not obvious if number *per se* or particular combinations is the key to success. Resistance genes can also be mixed within a field as a multiline cultivar or a cultivar mixture. Despite many experimental studies and a significant number of successes, it has been difficult to determine the types of pathosystem for which diversity provides a large degree of disease control. Still an issue is whether the use of mixtures will select complex races (super-races) that can overcome many or all of the resistance genes in the mixture, or will a stable equilibrium develop among races of varying complexity? Finally there is still much to learn about the effects of landscape-level processes on the population biology of plant pathogens and disease spread, and hence on whether or not landscapes can be managed to aid durability. Transgenic approaches may provide new opportunities for the control of some recalcitrant pathogens, but issues of durability for transgenes are likely to be the same as for other genes for resistance.

I will finish my account of disease and pest resistance as I started it, durability is still difficult to predict with certainty and there have been mixed fortunes in the past; but the need for it is as great as ever during a period of human population increase and climate change. More generally, the need for the genetic improvement of cultivated plants is greater than ever. But genetic improvement means adaptation to a farming system and end user needs, and probably the most difficult decision facing the next generation of plant breeders is the one I haven't answered, at least not directly: what are going to be the farming systems in 2050?

## Farming and the Fate of Wild Nature

I am therefore going to finish this final chapter with reference to a paper by Green et al. (2005) in *Science* which had the title "Farming and the Fate of Wild Nature". It addressed decisions about how to meet the world demand for food in 2050 and

their consequences for wild species and habitats. They showed that farming was already the greatest extinction threat to birds (the best known taxon), and that its adverse impacts looked set to increase, especially in developing countries. They then considered two competing solutions: (1) wildlife-friendly farming which boosts densities of wild populations on farmland but may decrease agricultural yields and (2) land sparing farming which minimizes demand for farmland by increasing yield. Finally they presented a model that attempted to resolve the trade-off between the two approaches. It showed that the best type of farming for species persistence depends on the demand for agricultural products and on how the population densities of different species on farmland change with agricultural yield. Evidence from a range of taxa in developing countries suggested that high-yield farming may allow more species to persist. Readers can look at the paper and draw their own conclusions. What interested me were the two visions of how and where we farm. In the first, there is a greater area of low-yielding wildlife-friendly farmland and less intact ‘natural’ habitat. In the second there is a smaller area of high-yielding, less wildlife-friendly farmland and more area available for wild nature elsewhere. Different decisions may be reached in different countries or even within countries. Where land sparing farming is adopted, experience has shown that relatively large short term gains have come from intensive farming of monoculture crops, often with relatively few cultivars grown over large areas. Yet in the longer term sustained high levels of overall production might come from greater diversity of crops and greater diversity of cultivars in regions, on individual farms, and even within individual fields. Whatever happens, there is going to be a need for as many new cultivars as possible that are suited to the diverse farming systems, environments and end user needs that continue to evolve up to and beyond 2050. Continued breeding of cultivars with the highest possible yields and quality for given inputs, combined with the conservation and evaluation of germplasm for later use, is the best insurance policy for an uncertain future (Fig. 18.6).



**Fig. 18.6** Two farming systems: intensive farming of monoculture crops and diversity of cultivars within individual fields. (a) A potato crop (cultivar Markies) in East Lothian, Scotland (1 July 2014). (b) A potato crop (many cultivars) near Lijiang, China (10 June 2014)

# **Epilogue**

## **Way Ahead**

### **Introduction**

The challenge facing plant breeders starting work in 2015 is to produce new cultivars of humankind's cultivated crops that will meet the needs of a global population estimated to reach nine billion in 2050; in particular to make sure that all of these people can be adequately fed in a sustainable and equitable manner. Current breeding goals in a wide range of crops can be found in the recently published volumes (1–10 so far) in the *Handbook of Plant Breeding* (Prohens et al. 2008). The goals are set in terms of producing new cultivars better adapted to the farming and production system under consideration, including levels of inputs such as fertilizer and water, and to the available growing season(s). They are also set in terms of meeting end user specifications. Lastly, consideration is given to the desirability and feasibility of inbuilt resistance to major pests and diseases, suppression of weeds, and tolerance of abiotic stresses such as heat, drought and salinity. Here in this epilogue are a few final thoughts on the bigger global picture in terms of timescales, scientific plant breeding and farming systems and new cultivars.

### **Timescales**

One hundred and fifteen years of scientific plant breeding has demonstrated that a lot can be achieved in a period of 35 years. Let us recall a few examples.

My first example is exploitation of heterosis in hybrid maize (*Zea mays*). Jones (1918) made his first double-cross hybrid in 1917; this was grown commercially in 1921; and then the period 1933 to 1950 saw an increase in double-cross hybrids in the USA from less than 1 % of the acreage to over 99 %, a remarkable achievement. Then over a second 30 year period from 1960, superior single-cross hybrids

replaced the double-cross ones and by 1990 accounted for over 85 % of production in the USA, thus making Shull's (Shull 1908) original research of 1908 on inbreeding and crossbreeding a commercial reality.

Oilseed rape (*Brassica napus*) provides an example of the speed and effectiveness of selection for altered chemical composition. Over the course of the last 30 years, starting in Canada, this temperate crop was transformed into the world's third most important source of both vegetable oil and animal meal. In the 1960s and 1970s government researchers in Canada bred a new type of rapeseed, canola, whose products were low erucic acid oil (<1 % of fatty acids compared with original 45 %) and low glucosinolates meal (<18 µmol g<sup>-1</sup> seed at 8.5 % moisture), thus making them suitable for use as edible and cooking oils, and animal meal, respectively. New analytical and breeding methods were the keys to success (Khachatourians et al. 2001). *B. napus* is tolerant of inbreeding and isolated microspore culture from F<sub>1</sub> plants proved an efficient way of producing doubled haploid lines and cultivars (Friedt and Snowdon 2009). Homozygous doubled haploid lines were also suitable for use in hybrid breeding, based on their *per se* performance and combining ability. Hybrid cultivars of winter rapeseed and spring canola have rapidly gained in importance since the year 2000 as effective systems were developed for controlled pollination. Yield improvements in European winter rapeseed of up to 15 % have been reported for F<sub>1</sub> hybrids compared to non-hybrids (Friedt and Snowdon 2009). Sunflower (*Helianthus annuus*) was also transformed into a major oilseed crop in the second half of the twentieth century due to a dramatic increase of oil percentage and the development of hybrid cultivars (Fernández-Martínez et al. 2009).

The new wheat cultivars of the "Green Revolution" were bred in Mexico and then India over the 30 year period from 1945 to 1974 and demonstrate the importance of selection environment and germplasm, as well as breeding method. The three key components to Borlaug's programme were use of large numbers of hybridizations combined with a search for stem rust resistance; shuttle breeding between the north and the highlands of Mexico to achieve two generations a year and a crop insensitive to day length; and use of short-strawed wheat, particularly Norin 10, which came to the USA from Japan after the Second World War. Similar progress was made with rice in the Philippines and Asia over the 30 year period from 1948, the year 1966 seeing the release of IR8. Rice hybrids were first produced in China in 1973, released in 1976, and by 1999 occupied over half the rice acreage. They have out yielded conventional inbred line cultivars by 20 % on average.

The production of potato cultivars suitable for growing as a winter crop in short days in the lowland subtropics (the Indo-Gangetic plains and southern China) was the result of major breeding efforts during the second half of the twentieth century, particularly in India since 1948 and in China since 1978. It is an example of expanding the area of cultivation of a crop through adaptation to new environments. Unlike maize there were no hybrids in which to exploit heterosis, and in contrast to wheat and rice, potatoes already had a high harvest index of around 0.80 (the proportion of the whole plant's dry-weight which is harvestable tuber). Adaptation

to the lowland subtropics can be seen as the final stage of the global expansion of potato growing, which initially was from the short summer days of the highland tropics and subtropics of the Andes to the long summer days of the lowland temperate regions of the world; an expansion that involved selection for the ability to form tubers in long days.

My last example is the rapid development and uptake, at least in some parts of the world, of genetically modified crops. Procedures for *Agrobacterium*-mediated and particle bombardment transformation were developed from 1984 and 1987, respectively. The first genetically modified (GM) crops were planted in 1996 and by 2014 occupied 181.5 million hectares of land in 28 countries, which is over 10 % of all cropland in the world (James 2014). This is clearly a remarkable achievement in a period of 30 years, albeit built on a previous 30 year period of rapid developments in molecular biology following the publication in 1953 of the molecular structure of DNA, the chemical carrier of genetic information (Watson and Crick 1953a, b).

It is worth reflecting on the fact that a lot was achieved during 115 years of scientific plant breeding because breeders were not starting from scratch with respect to their raw material (germplasm). Human selection of plants had begun when plants were first brought into cultivation for human use, as early as 13,000 years ago in the Near East, and subsequently and independently in many different parts of the world (Balter 2007). The plant species available for cultivation were the products of millions of years of evolution by natural selection.

The biochemistry of plant life was cracked some 2.7 billion years ago by the cyanobacteria, in the form of photosynthetic cells capable of using light energy from the sun, water, and carbon dioxide to synthesize complex sugars (with the release of oxygen).

The anatomy and physiology of plant life on dry land and in air was eventually completed with seeds which were fully enclosed within a specialized plant organ, the carpel. The result was the angiosperms whose rapid radiation some 100 million years ago left them the dominant part of the world flora. Their distribution 13,000 years ago was the result of the ensuing physical changes to planet earth.

Our present-day world took shape over the past 66 million years: the continental plates moved into their present positions, today's prominent mountain ranges formed, and new ocean current systems were established. An initially warm ( $>30^{\circ}\text{C}$ ) global climate became increasingly cool (towards  $15^{\circ}\text{C}$  global mean) and arid, culminating in a build-up of ice at the poles, greater temperature gradients from equator to poles, and the formation of deserts. Carbon dioxide concentrations fell from about 1200 ppm to around 300 ppm. All of these physical changes to the world resulted in major changes in the overall composition and distribution of its vegetation. For example, by around 20 million years ago in many areas, the grasses had become the dominant group of plants in truly open habitat grasslands; ecosystems that we now describe as tropical and subtropical savannah, temperate grasslands and steppes.

From about 2.5 million years ago, recurrent glacial-interglacial cycles occurred, closely linked to variations in the incoming solar radiation. Concentrations of atmospheric carbon dioxide fell to averages of around 180 ppm during the glacial

periods compared with 300 ppm during interglacial periods. During glacial periods temperate vegetation became isolated in regions where micro-environmentally favourable conditions existed, and then re-expanded from these regions during interglacial periods.

The germplasm available for scientific plant breeding was the result of farmer selection, initially over thousands of years and later over a few hundred years, as appreciated by Vavilov. The transition from exclusive use of wild foods to dependence on cultigens took about 2500 years and the average time span for the domestication syndrome of traits to be fixed in crops was 3767 years for trees and 2638 years for annuals (Meyer et al. 2012). Likewise the initial dispersion of crops like wheat and barley from their centres of origin took several thousand years; for example, 7000 years for the journey from the “Fertile Crescent” ( $33^{\circ}\text{N}$   $44^{\circ}\text{E}$ ) to what became their most north westward areas of production ( $57^{\circ}\text{N}$   $2^{\circ}\text{W}$ ). In contrast, the 500 years since Columbus discovered the New World in 1492 have seen an extensive global redistribution of many crops; for example, maize and potato from the New to the Old World and their establishment as two of the world’s most important food crops. The outcome of these dispersions and farmer selection was thousands of locally adapted landraces of cultivated plants which could be grouped into geographical races and ecotypes.

One result of scientific plant breeding has been the replacement of numerous landraces with relatively few high yielding cultivars. Furthermore, land degradation and changes in use have endangered the habitats of many of their wild relatives. Despite the superiority of modern cultivars in many present day farming systems, land races and wild relatives contain genetic variants that are desirable for use in future breeding and which have not been used to date by farmers and plant breeders. This novel genetic diversity will be needed to increase crop productivity and quality, disease and pest resistance and tolerance of adverse or marginal environments. Hence there is a need for both the *in situ* and *ex situ* conservation of this genetic diversity. Collection and storage of genetic diversity is not an end in itself, but rather the prerequisite for evaluation and use in plant breeding. Choosing the right germplasm will remain a challenge for plant breeders.

## Scientific Plant Breeding

The development of scientific plant breeding was based on an understanding of the mating systems of crop plants and the mechanism of inheritance, the legacies of Darwin and Mendel, respectively. In other words, an understanding of the genetic makeup of the gametes produced during meiosis (new combinations of alleles) and how these gametes are combined through the mating system to produce offspring with new genotypes, and hence desired phenotypes (traits). Effective plant breeding also required the ability to recognize and hence select these new genotypes. The twentieth century therefore saw the production of high yielding cultivars by appropriate artificial hybridization and selection: clonal cultivars from multistage,

multitrait selection; hybrid cultivars from inbreeding and crossbreeding; and inbred line cultivars from hybridization and inbreeding. Continued progress was made by cycles of hybridization and selection, usually among the developing elite germplasm. Sometimes new traits were required for new environments, new end uses, and new pest and disease problems, and sometimes broader genetic bases were sought to deal with perceived plateaus in progress. Thus the twentieth century saw limited use of wild relatives in introgression breeding and landraces in base broadening, as well as limited use of mutation breeding for special purposes. Finally, from the late 1970s breeders started to benefit in their selection programmes from the exciting developments in molecular biology, including the use of molecular markers to assist introgression breeding and to dissect complex traits such as yield. Breeders currently face the challenge of how best to prioritize and integrate DNA selection of desired alleles for qualitative traits and ones of large effect at QTLs (Quantitative Trait Loci) with field selection for quantitative traits; having first ensured that they are present among the parental germplasm. A new challenge is if and when to integrate genomic selection of quantitative traits. As a training population needs to be phenotyped as well as genotyped, a prerequisite may be the high throughput phenotyping methods currently being developed for use in the field.

Further improvements to successful cultivars can be made by genetic transformation and the production of genetically modified (GM) crops, with marker-free transformants and the ability to stack combinations of transgenes already realities. Furthermore, site-directed DNA sequence modification is now also becoming a reality; witness for example the current excitement over the CRISPR/Cas9 system for plant genome editing. The future scope of GM crops will increase with more detailed understanding of the biochemistry of a greater number of economically important traits.

The pace of technological advance quickened from 1995 with miniaturized microarrays for gene expression profiling of thousands of genes in parallel, and from 2004 with “Next-Generation Sequencing (NGS)” methods which dramatically increased the speed at which genomes could be sequenced. The result has been an equally rapid increase in knowledge of the genomes of cultivated plants and their evolutionary histories. Breeders starting work in 2015 will no doubt benefit from this increased genetic knowledge and combine it with the technological advances to aid the discovery of desirable genes (alleles) and to make breeding faster, more efficient and more effective at achieving desired goals. But technology alone will not be sufficient to meet the challenges of 2050. Breeders, as always, will need to apply appropriate breeding methods to the right germplasm for the right objectives, but the latter will depend upon answering the big questions about the most appropriate farming systems and most appropriate uses of crops for 2050.

## Farming Systems and New Cultivars

The history of plant breeding has been intimately associated with bringing land into cultivation and increasing its productivity. Indeed, one could say that plant breeding is the genetic improvement of cultivated plants by producing cultivars better adapted to particular farming systems for particular end uses.

The cultivation of crops has benefited from scientific improvements in farming, particularly since 1600. In the east of England, near Cambridge where I grew up, more land was brought into cultivation from 1600 to 1900 by the drainage, clearance and enclosure of 7000 km<sup>2</sup> of wet fens (Rotherham 2013). The period from 1600 to the present day also saw major changes in farming practice. The most fertile soils were used for arable crops, initially with a three-course rotation of 2 years of cereal crops followed by one of fallow. During the eighteenth century this was replaced by a four-course rotation of wheat, turnips, barley and clover. The turnips provided a break crop between cereals, a cleaning crop (from weeds) and good fodder for wintering livestock. The clover provided a nitrogen-fixing forage crop that left nitrogen in the soil for the subsequent cereal crops. Then the twentieth century saw the advent of new farming methods that used chemical fertilizers and selective herbicides. These changes removed the cultural need in arable systems for both the traditional break crop and the nitrogen-fixing crop, which could therefore be replaced by more intensive cereal growing and cash crops such as sugar beet, and later oilseed rape. The period from the 1950s has seen the development of intensive, petrochemically driven arable farming; an aspect of modern industrial farming which has become the norm in developed countries. Intensive farming of monoculture crops, often with relatively few cultivars grown over large areas, has given relatively large short term gains in production; but is this kind of farming sustainable and is it desirable in developing regions of the world? In the longer term sustained high levels of overall production might be achieved from greater diversity of crops and greater diversity of cultivars in regions, on individual farms, and even within individual fields. The FAO, IFAD and WFP Report (2015) on “The State of Food Insecurity in the World 2015” reminds us that more than 90 % of the 570 million farms worldwide are managed by an individual or a family, relying predominantly on family labour, and that 84 % of family farms are smaller than 2 ha. The amount of food production per hectare does not depend on the size of farm, but on the levels of inputs to meet the crops’ needs, which are relatively simple. Plants need light, water (H<sub>2</sub>O) and carbon dioxide (CO<sub>2</sub>) for photosynthesis. They also need to extract macro- and micro-nutrients (inorganic minerals) from the soil as ions in aqueous solution, unless grown directly in aqueous solutions of mineral nutrients (hydroponics). Hence the issue is whether or not the farming systems in 2050 will be able to provide the necessary inputs for the required levels of crop production.

Today about 12 % of the planet’s land mass, or 1.6 billion hectares, is cultivated for agriculture. The International Institute for Applied Systems Analysis (IIASA) has estimated that if protected areas such as national parks and also forests are

excluded, we have 1.3 billion hectares of grassland and open woodland suitable for agricultural expansion, of which 60 % is in South America and Africa. However, if possible, it would be better to get more food out of existing farmland; IIASA and the United Nations' FAO expect this to happen and therefore predict that the amount of cultivated land in the world will only grow by 10 % between now and 2050.

So who will make and implement the big policy decisions about increasing food production? Will it be world leaders through the United Nations, multinational companies, National Governments or simply the combined and independent actions of millions of farmers worldwide who want to feed their families and to supply their customers with desired farm produce? Whatever happens, it is the farmers who will need to grow as many new cultivars as possible that are suited to their diverse farming systems, environments and end user needs, all of which will continue to evolve up to and beyond 2050. Continued breeding of cultivars with the highest possible yields and quality for given inputs, combined with the conservation and evaluation of germplasm for later use, is the best insurance policy for an uncertain future. But there does need to be a sense of urgency about 2050 and an appreciation of the scale of breeding required.



# References

- Abberton MT, Marshall AH (2010) White clover. In: Boller B, Posselt UK, Veronesi F (eds) Fodder crops and amenity grasses, vol 5, Handbook of plant breeding. Springer, New York, pp 457–476
- Abdel-Ghani AH, Parzies HK, Omary A (2004) Estimating the outcrossing rate of barley landraces and wild barley populations collected from ecologically different regions of Jordan. *Theor Appl Genet* 109:588–595
- Abe A, Kosugi S, Yoshida K, Natsume S, Takagi H, Kanzaki H, Matsumura H, Yoshida K, Mitsuoka C, Tamiru M, Innan H, Cano L, Kamoun S, Terauchi R (2012) Genome sequencing reveals agronomically important loci in rice using MutMap. *Nat Biotechnol* 30:174–179
- Ahmad R, Khan RS, Farhatullah (2014) Transcript analyses of the gene associated with CMS in *Brassica napus*. *Euphytica* 200:455–463. doi:[10.1007/s10681-014-1195-0](https://doi.org/10.1007/s10681-014-1195-0)
- Allaby R (2010) Integrating the processes in the evolutionary system of domestication. *J Exp Bot* 61:935–944. doi:[10.1093/jxb/erp382](https://doi.org/10.1093/jxb/erp382)
- Allaby M (2012) Oxford dictionary of plant sciences, 3rd edn. Oxford University Press, Oxford, 565p
- Allard RW (1999) Principles of plant breeding, 2nd edn. Wiley, New York, 254p
- Allman ES, Rhodes JA (2004) Mathematical models in biology. Cambridge University Press, Cambridge, 370p
- Alvarez N, Garine E, Khasah C, Dounias E, Hossaert-McKey M, McKey D (2005) Farmers' practices, metapopulation dynamics, and conservation of agricultural biodiversity on-farm: a case study of sorghum among the Duupa in sub-sahelian Cameroon. *Biol Conserv* 121:533–543
- Ames M, Spooner DA (2008) DNA from herbarium specimens settles a controversy about origins of the European potato. *Am J Bot* 95:252–257
- Amuzu-Aweh EN, Bijma P, Kinghorn BP, Vereijken A, Visscher J, van Arendonk JAM, Bovenhuis H (2013) Prediction of heterosis using genome-wide SNP-marker data: application to egg production traits in white Leghorn crosses. *Heredity* 111:530–538. doi:[10.1038/hdy.2013.77](https://doi.org/10.1038/hdy.2013.77)
- Anderson T (1855) Proceedings in the laboratory. Transactions of the Highland and Agricultural Society of Scotland for 1855 to 1857, pp 174–200
- Andrivon D, Pellé R, Ellissèche D (2006) Assessing resistance types and levels to epidemic diseases from the analysis of disease progress curves: principles and application to potato late blight. *Am J Potato Res* 83:455–461

- Angioi SA, Rau D, Attene G, Nanni L, Bellucci E, Logozzo G, Negri V, Spagnoletti Zeuli PL, Papa R (2010) Beans in Europe: origin and structure of the European landraces of *Phaseolus vulgaris* L. *Theor Appl Genet* 121:829–843
- Araus JL, Cairns JE (2014) Field high-throughput phenotyping: the new crop breeding frontier. *Trends Plant Sci* 19:52–61. doi:[10.1016/j.tplants.2013.09.008](https://doi.org/10.1016/j.tplants.2013.09.008)
- Arnau G, Abraham K, Sheela MN, Chair H, Sartie A, Asiedu R (2010) Yams. In: Bradshaw JE (ed) Root and tuber crops, vol 7, Handbook of plant breeding. Springer, New York, pp 127–148
- Asano K, Takashi T, Miura K, Qian Q, Kitano H, Matsuoka M, Ashikari M (2007) Genetic and molecular analysis of utility of *sdl* alleles in rice breeding. *Breed Sci* 57:53–58
- Atlin GN, Cooper M, Bjørnstad Å (2001) A comparison of formal and participatory breeding approaches using selection theory. *Euphytica* 122:463–475
- Avery OT, Macleod CM, McCarty M (1944) Studies on the chemical nature of the substance inducing transformation of pneumococcal types. Induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus Type III. *J Exp Med* 79:137–158
- Azzalini A, Cox DR (1984) Two new tests associated with analysis of variance. *J R Stat Soc Ser B* 46:335–343
- Badaeva ED, Dedkova OS, Gay G, Pukhalskyi VA, Zelenin AV, Bernard S, Bernard M (2007) Chromosomal rearrangements in wheat: their types and distribution. *Genome* 50:907–926
- Badenes ML, Byrne DH (eds) (2012) Fruit breeding, vol 8, Handbook of plant breeding. Springer, New York, 875p
- Baenziger P, Graybosch R, Van Sanford D, Berzonsky W (2009) Winter and specialty wheat. In: Carena MJ (ed) Cereals, vol 3, Handbook of plant breeding. Springer, New York, pp 251–265
- Bai Y (2012) Introgression breeding on tomatoes for resistance to powdery mildew. In: Acquaah G (ed) Principles of plant genetics and breeding. Wiley-Blackwell, Chichester, pp 55–60
- Bai Y, Lindhout P (2007) Domestication and breeding of tomatoes: what have we gained and what can we gain in the future? *Ann Bot* 100:1085–1094. doi:[10.1093/aob/mcm150](https://doi.org/10.1093/aob/mcm150)
- Bairu MW, Aremu AO, Van Staden J (2011) Somaclonal variation in plants: causes and detection methods. *Plant Growth Regul* 63:147–173
- Bajer A (1957–1958) Ciné-micrographic studies on mitosis in endosperm. *Exp Cell Res* 13: 493–502, 14: 245–256, 15: 370–383
- Baker RJ (1988) Tests for crossover genotype-environmental interactions. *Can J Plant Sci* 68:405–410
- Baldoni L, Belaj A (2009) Olives. In: Vollmann J, Rajcan I (eds) Oil crops, vol 4, Handbook of plant breeding. Springer, New York, pp 397–421
- Balestre M, Von Pinho RG, Souza JC, Oliveira RL (2009) Potential use of molecular markers for prediction of genotypic values in hybrid maize performance. *Genet Mol Res* 8:1292–1306
- Balter M (2007) Seeking agriculture's ancient roots. *Science* 316:1830–1835
- Banga O (1957a) Origin of the European cultivated carrot. *Euphytica* 6:54–63
- Banga O (1957b) The development of the original European carrot material. *Euphytica* 6:64–76
- Bannerot ML, Louidard Y, Cauderon Y, Tempe J (1974) Transfer of cytoplasmic male sterility from *Raphanus sativus* to *Brassica oleracea*. In: Proceedings of Eucarpia Meeting Cruciferae, Dundee, pp 52–54
- Barnaud A, Trigueros G, McKey D, Joly HI (2008) High outcrossing rates in fields with mixed sorghum landraces: how are landraces maintained? *Heredity* 101:445–452. doi:[10.1038/hdy.2008.77](https://doi.org/10.1038/hdy.2008.77)
- Barone A, Sebastian A, Carpoto D, della Rocca F, Frusciante L (2001) Molecular marker-assisted introgression of the wild *Solanum commersonii* genome into the cultivated *S. tuberosum* gene pool. *Theor Appl Genet* 102:900–907
- Barrett SCH (ed) (2008) Major evolutionary transitions in flowering plant reproduction. The University of Chicago Press, Chicago, 209p
- Barry TN (2013) The feeding value of forage brassica plants for grazing ruminant livestock. *Anim Feed Sci Technol* 181:15–25

- Bar-Yosef O (1998) The Natufian culture in the Levant, threshold to the origins of agriculture. *Evol Anthropol* 6:159–177
- Bateson W (1907) Facts limiting the theory of heredity. *Science* 26:649–660
- Bateson W, Saunders ER, Punnett RC (1905) Experimental studies in the physiology of heredity. *Rep Evol Comm R Soc* 2:1–55, 80–99
- Batista R, Saibo N, Lourenco T, Oliveira MM (2008) Microarray analyses reveal that plant mutagenesis may induce more transcriptomic changes than transgene insertion. *Proc Natl Acad Sci* 105:3640–3645
- Bauman LF (1981) Review of methods used by breeders to develop superior corn inbreds. *Proc Annu Corn Sorghum Ind Res Conf* 36:199–208
- Beadle GW (1945) Genetics and metabolism in *Neurospora*. *Physiol Rev* 25:643–663
- Beadle GW, Tatum EL (1941) Genetic control of biochemical reactions in *Neurospora*. *Proc Natl Acad Sci* 27:499–506
- Becker HC (1988) Breeding synthetic varieties of crop plants. *Plant Genet Breed Rev* 1:31–54
- Becker HC, Damgaard C, Karlsson B (1992) Environmental variation for outcrossing rate in rapeseed. *Theor Appl Genet* 84:303–306
- Beckett JB (1971) Classification of male-sterile cytoplasms in maize (*Zea mays* L.). *Crop Sci* 11:724–727
- Beilstein MA, Nagalinguma NS, Clements MD, Manchester SR, Mathewson S (2010) Dated molecular phylogenies indicate a Miocene origin for *Arabidopsis thaliana*. *Proc Natl Acad Sci* 107:18724–18728
- Bell A (2009) Peak water. Luath, Edinburgh, 207p
- Belling J (1928) A working hypothesis for segmental interchange between homologous chromosomes in flowering plants. *Univ Calif Publ Bot* 14:283–291
- Bennett JH, Binet FE (1956) Association between Mendelian factors with mixed selfing and random mating. *Heredity* 10:51–55
- Benson W (2012) Kingdom of plants. A journey through their evolution. Collins, London, 256p
- Bernardo R (1991) Correlation between testcross performance of lines at early and late selfing generations. *Theor Appl Genet* 82:17–21
- Berners-Lee M, Clark D (2013) The burning question. Profile Books, London, 268p
- Bernoux M, Ve T, Williams S, Warren C, Hatters D, Valkov E, Zhang X, Ellis JG, Kobe B, Dodds PN (2011) Structural and functional analysis of a plant resistance protein TIR domain reveals interfaces for self-association, signaling, and autoregulation. *Cell Host Microbe* 9:200–211
- Betrán FJ (2012) Hybrid breeding in maize. In: Acquaah G (ed) Principles of plant genetics and breeding. Wiley-Blackwell, Chichester, pp 597–601
- Bhattacharyya MK, Smith AM, Ellis TH, Hedley C, Martin C (1990) The wrinkled-seed character of pea described by Mendel is caused by a transposon-like insertion in a gene encoding starch-branched enzyme. *Cell* 60:115–122
- Biancardi E, McGrath JM, Panella LW, Lewellen RT, Stevanato P (2010) Sugar beet. In: Bradshaw JE (ed) Root and tuber crops, vol 7, Handbook of plant breeding. Springer, New York, pp 173–219
- Biffen RH (1905) Mendel's laws of inheritance and wheat breeding. *J Agric Sci* 1:4–48
- Biffen RH (1912) Studies in inheritance in disease resistance II. *J Agric Sci* 4:421–429
- Bingham J (1979) Wheat breeding objectives and prospects. *Agric Prog* 54:1–17
- Bingham ET, McCoy TJ (1988) Cytology and cytogenetics of alfalfa. In: Hanson AA (ed) Alfalfa and alfalfa improvement, Agronomy monographs 29. ASA, CSSA, SSSA, Madison, pp 737–776
- Binglun Z (2009) Researches in heredity and breeding. In: Ancient China's technology and science. Foreign Languages Press, Beijing, pp 281–291
- Bitocchi E, Nanni L, Bellucci E, Rossi M, Giardini A, Zeuli PS, Logozzo G, Stougaard J, McClean P, Attene G, Papa R (2012) Mesoamerican origin of the common bean (*Phaseolus vulgaris* L.) is revealed by sequence data. *Proc Natl Acad Sci* 109:E788–E796. doi:[10.1073/pnas.1108973109](https://doi.org/10.1073/pnas.1108973109)

- Bitocchi E, Bellucci E, Giardini A, Rau D, Rodriguez M, Biagetti E, Santilocchi R, Zeuli PS, Gioia T, Logozzo G, Attene G, Nanni L, Papa R (2013) Molecular analysis of the parallel domestication of the common bean (*Phaseolus vulgaris*) in Mesoamerica and the Andes. *New Phytol* 197:300–313. doi:[10.1111/j.1469-8137.2012.04377.x](https://doi.org/10.1111/j.1469-8137.2012.04377.x)
- Black W, Mastenbroek C, Mills WR, Peterson LC (1953) A proposal for an international nomenclature of races of *Phytophthora infestans* and of genes controlling immunity in *Solanum demissum* derivatives. *Euphytica* 2:173–179
- Blackmore S (2012) Green Universe. Royal Botanic Garden Edinburgh/Papadakis, Winterbourne, 256 p
- Blakeslee AF, Avery AG (1937) Methods of inducing doubling of chromosomes in plants, by treatment with colchicine. *J Hered* 28:392–411
- Blum A (2005) Drought resistance, water-use efficiency, and yield potential—are they compatible, dissonant, or mutually exclusive? *Aust J Agric Res* 56:1159–1168
- Blum A (2014) Genomics for drought resistance—getting down to earth. *Funct Plant Biol* 41:1191–1198. doi:[10.1071/FP14018](https://doi.org/10.1071/FP14018)
- Bohn M, Utz HF, Melchinger AE (1999) Genetic similarities among winter wheat cultivars determined on the basis of RFLPs, AFLPs, and SSRs and their use for predicting progeny variance. *Crop Sci* 39:228–237
- Boller B, Greene SL (2010) Genetic resources. In: Boller B, Posselt UK, Veronesi F (eds) *Fodder crops and amenity grasses*, vol 5, *Handbook of plant breeding*. Springer, New York, pp 13–37
- Boller B, Posselt UK, Veronesi F (eds) (2010a) *Fodder crops and amenity grasses*, vol 5, *Handbook of plant breeding*. Springer, New York, 523p
- Boller B, Schubiger FX, Kölliker R (2010b) Red clover. In: Boller B, Posselt UK, Veronesi F (eds) *Fodder crops and amenity grasses*, vol 5, *Handbook of plant breeding*. Springer, New York, pp 439–455
- Bortesi L, Fischer R (2015) The CRISPR/Cas9 system for plant genome editing and beyond. *Biotechnol Adv* 33:41–52
- Bradshaw JE (1983) Estimating the predicted response to  $S_1$  family selection. *Heredity* 51:415–418
- Bradshaw JE (1984) Computer simulation of family selection schemes suitable for kale (*Brassica oleracea* L.), involving half-sib, full-sib and selfed families. *Theor Appl Genet* 68:503–508
- Bradshaw JE (1986) Competition between cultivars of fodder kale in yield trials with single-row plots. *Euphytica* 35:433–439
- Bradshaw JE (1987) The choice of selection index in kale (*Brassica oleracea* L.) population improvement. *Theor Appl Genet* 75:165–169
- Bradshaw JE (1989) Interplot competition in yield trials of swedes. *Euphytica* 42:135–140
- Bradshaw JE (1994b) Quantitative genetics theory for tetrasomic inheritance. In: Bradshaw JE, Mackay GR (eds) *Potato genetics*. CAB International, Wallingford, pp 71–99
- Bradshaw JE (2007) Breeding potato as a major staple crop. In: Kang MS, Priyadarshan PM (eds) *Breeding major food staples*. Blackwell, Oxford, pp 277–332
- Bradshaw JE (2009a) Potato breeding at the Scottish Plant Breeding Station and the Scottish Crop Research Institute: 1920–2008. *Potato Res* 52:141–172
- Bradshaw JE (2009b) A genetic perspective on yield plateau in potato. *Potato J* 36:79–94
- Bradshaw JE (2009c) Breeding for field resistance to late blight of potato at SCRI. *Acta Hortic* 834:87–100
- Bradshaw JE (ed) (2010) *Root and tuber crops*, vol 7, *Handbook of plant breeding*. Springer, New York, 295p
- Bradshaw JE, Bonierbale M (2010) Potatoes. In: Bradshaw JE (ed) *Root and tuber crops*, vol 7, *Handbook of plant breeding*. Springer, New York, pp 1–52
- Bradshaw JE, Mackay GR (1985) Half-sib family selection for yield of digestible organic matter in kale (*Brassica oleracea* L.). *Euphytica* 34:201–206
- Bradshaw JE, Mackay GR (1994a) Potato genetics. CAB International, Wallingford, 552p

- Bradshaw JE, Mackay GR (1994b) Breeding strategies for clonally propagated potatoes. In: Bradshaw JE, Mackay GR (eds) Potato genetics. CAB International, Wallingford, pp 467–497
- Bradshaw JE, Ramsay G (2005) Utilisation of the Commonwealth Potato Collection in potato breeding. *Euphytica* 146:9–19
- Bradshaw JE, Wilson RN (1993) Inbred line versus F<sub>1</sub> hybrid breeding in swedes (*Brassica napus* L. var. *Napobrassica* Peterm). *Ann Appl Biol* 123:657–665
- Bradshaw JE, Wilson RN (2012) Kale population improvement and cultivar production. *Euphytica* 184:275–288
- Bradshaw JE, Dale MFB, Swan GEL, Todd D, Wilson RN (1998) Early-generation selection between and within pair crosses in a potato (*Solanum tuberosum* subsp. *tuberosum*) breeding programme. *Theor Appl Genet* 97:1331–1339
- Bradshaw JE, Lees AK, Stewart HE (2000) How to breed potatoes for resistance to fungal and bacterial diseases. *Plant Breed Seed Sci* 44:3–20
- Bradshaw JE, Pande B, Bryan GJ, Hackett CA, McLaren K, Stewart HE, Waugh R (2004) Interval mapping of quantitative trait loci for resistance to late blight [*Phytophthora infestans* (Mont.) de Bary], height and maturity in a tetraploid population of potato (*Solanum tuberosum* subsp. *tuberosum*). *Genetics* 168:983–995
- Bradshaw JE, Titley M, Wilson RN (2009a) Single seed descent as a breeding method for swedes (*Brassica napus* L. var. *napoibrassica* Peterm). *Euphytica* 169:387–401
- Bradshaw JE, Dale MFB, Mackay GR (2009b) Improving the yield, processing quality and disease and pest resistance of potatoes by genotypic recurrent selection. *Euphytica* 170:215–227
- Bregger T (1918) Linkage in maize: the C aleurone factor and waxy endosperm. *Am Nat* 52:57–61
- Briggs FN, Knowles PF (1967) Introduction to plant breeding. Reinhold, New York, 426p
- Britt AB (1996) DNA damage and repair in plants. *Annu Rev Plant Physiol Plant Mol Biol* 47:75–100
- Brookes G (2014) Weed control changes and genetically modified herbicide tolerant crops in the USA 1996–2012. *GM Crops Food* 5:321–332
- Brown R (1833) On the organs and mode of fecundation in Orchideae and Asclepiadaceae. *Trans Linn Soc London* 16:685–745
- Brown AHD (1989) The case for core collections. In: Brown AHD, Marshall DR, Frankel OH, Williams JT (eds) The use of plant genetic resources. Cambridge University Press, Cambridge, pp 136–156
- Brown CR (1993) Outcrossing rate in cultivated autotetraploid potato. *Am J Potato Res* 70:725–734
- Brown S (2012) Apple. In: Badenes ML, Byrne DH (eds) Fruit breeding, vol 8, Handbook of plant breeding. Springer, New York, pp 329–367
- Bruce AB (1910) The Mendelian theory of heredity and the augmentation of vigor. *Science* 32:627–628
- Bryan GJ, McLean K, Bradshaw JE, De Jong WS, Phillips M, Castelli L, Waugh R (2002) Mapping QTLs for resistance to the cyst nematode *Globodera pallida* derived from the wild potato species *Solanum vernei*. *Theor Appl Genet* 105:68–77
- Bryan GJ, McLean K, Pande B, Purvis A, Hackett CA, Bradshaw JE, Waugh R (2004) Genetical dissection of H3-mediated polygenic PCN resistance in a heterozygous autotetraploid potato population. *Mol Breed* 14:105–116
- Bulmer MG (1967) Principles of statistics, 2nd edn. Oliver & Boyd, Edinburgh, 252p
- Bulmer MG (1980) The mathematical theory of quantitative genetics. Clarendon, Oxford, 255p
- Butler D (2013) Fungus threatens top banana. *Nature* 504:195–196
- Byarugaba AA, Prossy N, Kashaija IN (2013) Identification of potato clones of population B3C2 with durable field resistance to late blight (*Phytophthora infestans*) and high yields in Uganda. *Afr J Agric Res* 8:3055–3059. doi:[10.5897/AJAR2013.6917](https://doi.org/10.5897/AJAR2013.6917)
- Caldwell DG, McCallum N, Shaw P, Muehlbauer GJ, Marshall DF, Waugh R (2004) A structured mutant population for forward and reverse genetics in barley (*Hordeum vulgare* L.). *Plant J* 40:143–150. doi:[10.1111/j.1365-313X.2004.02190.x](https://doi.org/10.1111/j.1365-313X.2004.02190.x)

- Callaway DJ, Callaway MB (eds) (2000) Breeding ornamental plants. Timber, Portland, 323p
- Camadro EL, Carputo D, Peloquin SJ (2004) Substitutes for genome differentiation in tuber-bearing *Solanum*: interspecific pollen-pistal incompatibility, nuclear-cytoplasmic male sterility, and endosperm. *Theor Appl Genet* 109:1369–1376
- Camadro EL, Erazzu LE, Maune JF, Bedogni MC (2012) A genetic approach to the species problem in wild potato. *Plant Biol* 14:543–554
- Cardi T, Earle ED (1997) Production of new CMS *Brassica oleracea* by transfer of ‘Anand’ cytoplasm from *B. rapa* through protoplast fusion. *Theor Appl Genet* 94:204–212
- Carlson EA (2004) Mendel’s legacy: the origin of classical genetics. Cold Spring Harbor Laboratory Press, New York, 332p
- Carlson PS, Smith HH, Dearing RD (1972) Parasexual interspecific plant hybridization. *Proc Natl Acad Sci* 69:2292–2294
- Caromel B, Mugnieri D, Kerlan M-C, Andrzejewski S, Palloix A, Ellisseche D, Rouselle-Bourgeois F, Lefebvre V (2005) Resistance quantitative trait loci originating from *Solanum sparsipilum* act independently on the sex ratio of *Globodera pallida* and together for developing a necrotic reaction. *Mol Plant Microbe Interact* 18:1186–1194
- Carputo D, Barone A (2005) Ploidy level manipulations in potato through sexual hybridisation. *Ann Appl Biol* 146:71–79
- Carputo D, Frusciante L (2011) Classical genetics and traditional breeding. In: Bradeen JM, Kole C (eds) Genetics, genomics and breeding of potato. Science, Enfield, pp 20–40
- Carson R (1962) Silent spring. Penguin Books, Harmondsworth, 317p
- Casler MD, Brummer EC (2008) Theoretical expected genetic gains for among-and-within-family selection methods in perennial forage crops. *Crop Sci* 48:890–902
- Castiglion P, Warner D, Bensen RJ, Anstrom DC, Harrison J, Stoecker M, Abad M, Kumar G, Salvador S, D’Ordine R, Navarro S, Back S, Fernandes M, Targoll J, Dasgupta S, Bonin C, Luethy MH, Heard JE (2008) Bacterial RNA chaperones confer abiotic stress tolerance in plants and improved grain yield in maize under water-limited conditions. *Plant Physiol* 147:455. doi:[10.1104/pp.108.118828](https://doi.org/10.1104/pp.108.118828)
- Causton DR (1977) A biologist’s mathematics. Edward Arnold, London, 326p
- Causton HC, Quackenbush J, Brazma A (2003) Microarray gene expression data analysis. Blackwell, Oxford, 160p
- Ceballos H, Okogbenin E, Pérez J, López-Valle L, Debouck D (2010) Cassava. In: Bradshaw JE (ed) Root and tuber crops, vol 7, Handbook of plant breeding. Springer, New York, pp 53–96
- Ceccarelli S, Grando S (2009) Participatory plant breeding in cereals. In: Carena MJ (ed) Cereals, vol 3, Handbook of plant breeding. Springer, New York, pp 395–414
- Chandler CK, Folta K, Dale A, Whitaker VM, Herrington M (2012) Strawberry. In: Badenes ML, Byrne DH (eds) Fruit breeding, vol 8, Handbook of plant breeding. Springer, New York, pp 305–325
- Chandra S, Huaman Z, Krishna H, Ortiz R (2002) Optimal sampling strategy and core collection size of the Andean tetraploid potato based on isozyme analysis data-a simulation study. *Theor Appl Genet* 104:1325–1334
- Chang MM, Culley D, Choi JJ, Hadwiger LA (2002) *Agrobacterium*-mediated co-transformation of a pea beta-1,3-glucanase and chitinase genes in potato (*Solanum tuberosum* L. cv. Russet Burbank) using a single selectable marker. *Plant Sci* 163:83–89
- Charlesworth B, Charlesworth D (2010) Elements of evolutionary genetics. Roberts, Greenwood Village, 734p
- Charlesworth D, Vekemans X, Castric V, Glémén S (2005) Plant self-incompatibility systems: a molecular evolutionary perspective. *New Phytol* 168:61–69
- Chen L, Liu YG (2014) Male sterility and fertility restoration in crops. *Annu Rev Plant Biol* 65:579–606
- Chester M, Leitch AR, Soltis PS, Soltis DE (2010) Review of the application of modern cytogenetic methods (FISH/GISH) to the study of reticulation (polyploidy/hybridisation). *Genes* 1:166–192. doi:[10.3390/genes1020166](https://doi.org/10.3390/genes1020166)

- Chetelat RT, Cisneros P, Stamova L, Rick CM (1997) A male-fertile *Lycopersicon esculentum*-*Solanum lycopersicoides* hybrid enables direct backcrossing to tomato at the diploid level. *Euphytica* 95:99–108
- Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative trait mapping. *Genetics* 138:963–971
- Cober ER, Cianzio SR, Pantalone VR, Rajcan I (2009) Soybean. In: Vollmann J, Rajcan I (eds) Oil crops, vol 4, Handbook of plant breeding. Springer, New York, pp 57–90
- Cochran WG (1951) Improvement by means of selection. In: Proceedings of the second Berkeley symposium in mathematical statistics and probability, Berkeley, 1951, pp 449–470
- Cockerham CC, Matzinger DF (1985) Selection response based on selfed progenies. *Crop Sci* 25:483–488
- Cockerham CC, Zeng Z-B (1996) Design III with marker loci. *Genetics* 143:1437–1456
- Cockram J, White J, Zuluaga DL, Smith D, Comadran J, Macaulay M, Luo Z, Kearsey MJ, Werner P, Harrap D, Tapsell C, Liu H, Hedley PE, Stein N, Schulte D, Steuernagel B, Marshall DF, Thomas WTB, Ramsay L, Mackay I, Balding DJ, The AGOUEB Consortium, Waugh R, O'Sullivan DM (2010) Genome-wide association mapping to candidate polymorphism resolution in the unsequenced barley genome. *Proc Natl Acad Sci* 107:21611–21616
- Cohen SN, Chang ACY, Hsu L (1972) Nonchromosomal antibiotic resistance in bacteria: genetic transformation of *Escherichia coli* by R-factor DNA. *Proc Natl Acad Sci* 69:2110–2114
- Collard BCY, Jahufer MZZ, Brouwer JB, Pang ECK (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. *Euphytica* 142:169–196
- Comstock RE, Robinson HF (1952) Estimation of average dominance of genes. In: Gowen JW (ed) Heterosis. Iowa State College Press, Ames, pp 494–516
- Comstock RE, Robinson HF, Harvey PH (1949) A breeding procedure designed to make maximum use of both general and specific combining ability. *Agron J* 41:360–367
- Connolly T, Currie ID, Bradshaw JE, McNicol JW (1993) Interplot competition in yield trials of potatoes (*Solanum tuberosum* L.) with single-drill plots. *Ann Appl Biol* 123:367–377
- Cooke DEL, Lees AK, Shaw DS, Taylor MC, Prentice MWC, Bradshaw NJ, Bain RA (2008) The status of GB blight populations and the threat of oospores. In: Proceedings Crop Protection in Northern Britain 2008, Dundee, pp 217–222
- Cooper HD, Spillane C, Hodgkin T (2001) Broadening the genetic base of crop production. CAB International, Wallingford, 452p
- Cooper JL, Till BJ, Laport RG, Darlow MC, Kleffner JM, Jamai A, El-Mellouki T, Liu S, Ritchie R, Nielsen N, Bilyeu KD, Meksem K, Comai L, Henikoff S (2008) TILLING to detect induced mutations in soybean. *BMC Plant Biol* 8:9. doi:[10.1186/1471-2229-8-9](https://doi.org/10.1186/1471-2229-8-9)
- Cornish MA (1990a) Selection during a selfing programme. I. The effects of a single round of selection. *Heredity* 65:201–211
- Cornish MA (1990b) Selection during a selfing programme. II. The effects of two or more rounds of selection. *Heredity* 65:213–220
- Cox MC, Hogarth DM (1993) The effectiveness of family selection in early stages of a sugarcane improvement program. In: Proceedings of Australian plant breeding conference, Bundaberg, April 1993, vol 10. Brisbane: Watson Ferguson, pp 53–54, ISBN 1864239573
- Craig WF (1977) Production of hybrid seed corn. In: Sprague GF (ed) Corn and corn improvement, vol 18, Agronomy., pp 671–719
- Craufurd PQ, Wheeler TR (2009) Climate change and the flowering time of annual crops. *J Exp Bot* 60:2529–2539
- Creighton HB, McClintock B (1931) A correlation of cytological and genetic crossing-over in *Zea mays*. *Proc Natl Acad Sci* 17:492–497
- Crossa J, de los Campos G, Perez P, Gianola D, Burgueno J, Araus JL, Makumbi D, Singh RP, Dreisigacker S, Yan J, Arief V, Banziger M, Braun H-J (2010) Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics* 186:713–724

- Crossa J, Pérez P, Hickey J, Burgueño J, Ornella L, Cerón-Rojas J, Zhang X, Dreisigacker S, Babu R, Li Y, Bonnett D, Mathews K (2014) Genomic prediction in CIMMYT maize and wheat breeding programs. *Heredity* 112:48–60. doi:[10.1038/hdy.2013.16](https://doi.org/10.1038/hdy.2013.16)
- Crow JF, Kimura M (1970) An introduction to population genetics theory. Harper & Row, New York, 591p
- Cruickshank G, Stewart HE, Wastie RL (1982) An illustrated assessment key for foliage blight of potatoes. *Potato Res* 25:213–214
- Cuadrado A, Acevedo R, DóÂaz de la Espina SM, Jouve N, de la Torre C (2004) Genome remodelling in three modern *S. officinatum* × *S. spontaneum* sugarcane cultivars. *J Exp Bot* 55:847–854. doi:[10.1093/jxb/erh093](https://doi.org/10.1093/jxb/erh093)
- Cuenca J, Froelicher Y, Aleza P, Juárez J, Navarro L, Ollitrault P (2011) Multilocus half-tetrad analysis and centromere mapping in citrus: evidence of SDR mechanism for  $2n$  megagametophyte production and partial chiasma interference in mandarin cv ‘Fortune’. *Heredity* 107:462–470
- Curtin SJ, Voytas DF, Stupar RM (2012) Genome engineering of crops with designer nucleases. *Plant Genome* 5:42–50
- Dale MFB, Phillips MS, Ayres RM, Hancock M, Holliday M, Mackay GR, Tones SJ (1988) The assessment of the tolerance of partially resistant potato clones to damage by the potato cyst nematode *Globodera pallida* at different sites and in different years. *Ann Appl Biol* 113:79–88
- Darlington CD (1934) The origin and behaviour of chiasmata. VII. *Zea mays*. *Z VererbLehre* 67:96–114
- Darrah LL, Zuber MS (1986) 1985 United States farm maize germplasm base and commercial breeding strategies. *Crop Sci* 26:1109–1113
- Darwin C (1859) On the origin of species by means of natural selection. The 150th Anniversary Landmark Edition: Bynum W (ed) Penguin Group, London, 516p
- Darwin C (1868) The variation of animals and plants under domestication, vol 1. Digitally printed 2010, Cambridge University Press, Cambridge, 411p
- Darwin C (1876) The effects of cross and self fertilisation in the vegetable kingdom. Digitally printed 2009, Cambridge University Press, Cambridge, 482p
- Darwin C (1877) The different forms of flowers on plants of the same species. Digitally printed 2010, Cambridge University Press, Cambridge, 352p
- Davenport CB (1908) Degeneration, albinism and inbreeding. *Science* 28:454–455
- Davies HV (2002) Commercial developments with transgenic potato. In: Valpuesta V (ed) *Fruit and vegetable biotechnology*. Woodhead, Cambridge, pp 222–249
- Dawson JC, Goldringer I (2012) Breeding for genetically diverse populations: variety mixtures and evolutionary populations. In: Lammerts van Buren ET, Myers JR (eds) *Organic crop breeding*. Wiley, Chichester, pp 77–98
- Dawson JC, Huggins DR, Jones SS (2008) Characterizing nitrogen use efficiency in natural and agricultural ecosystems to improve the performance of cereal crops in low-input and organic agricultural systems. *Field Crop Res* 107:89–101
- Day PR (1974) Genetics of host-parasite interaction. Freeman, San Francisco, 238p
- De Block M, Herrera-Estrella L, Van Montagu M, Schell J, Zambryski P (1984) Expression of foreign genes in regenerated plants and in their progeny. *EMBO J* 3:1681–1689
- de los Campos G, Hickey JM, Pong-Wong R, Daetwyler HD, Calus MPL (2013) Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics* 193:327–345
- Delourme R, Foisset N, Horvais R, Barret P (1998) Characterisation of the radish introgression carrying the *Rfo* restorer gene for the Ogu-INRA cytoplasmic male sterility in rapeseed (*Brassica napus* L.). *Theor Appl Genet* 97:129–134
- Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc Ser B* 39:1–38
- Desclaux D, Ceccarelli S, Navazio J, Coley M, Trouche G, Aguirre S, Weltzien E, Lançon J (2012) Centralized or decentralized breeding: the potentials of participatory approaches for low-input

- and organic agriculture. In: Lammerts van Bueren ET, Myers JR (eds) *Organic crop breeding*. Wiley, Chichester, pp 99–123
- Despeghel JP, Guguen N (2012) Feeding the future; breeding advances in oilseed rape oils. In: Bedő Z, Láng L (eds) *Plant breeding for future generations*, Proceedings of the 19th EUCARPIA General Congress, Agricultural Institute, Centre for Agricultural Research, Hungarian Academy of Sciences Martonvásár, Hungary, pp 62–65
- Devi J, Ko JM, Seo BB (2005) FISH and GISH: modern cytogenetic techniques. *Indian J Biotechnol* 4:307–315
- D'Hont A, Grivet L, Feldmann P, Rao S, Berding N, Glaszmann C (1996) Characterization of the double genome structure of modern sugarcane cultivars (*Saccharum* spp.) by molecular cytogenetics. *Mol Gen Genet* 250:405–413
- D'Hont A, Ison D, Alix K, Roux C, Glaszmann C (1998) Determination of basic chromosome numbers in the genus *Saccharum* by physical mapping of ribosomal RNA genes. *Genome* 41:221–225
- Díez MJ, Nuez F (2008) Tomato. In: Prohens J, Nuez F (eds) *Vegetables II*, vol 2, *Handbook of plant breeding*. Springer, New York, pp 249–323
- Dikötter F (2010) Mao's great famine. Bloomsbury, London, 420p
- Dixon GR (2007) Vegetable Brassicas and related Crucifers. *Crop production science in horticulture*, vol 14. CABI, Wallingford, 327p
- Dodds KS (1962) Classification of cultivated potatoes. In: Correll DS (ed) *The potato and its wild relatives*. Texas Research Foundation, Renner, pp 517–539
- Dodds KS (1965) The history and relationships of cultivated potatoes. In: Hutchinson JB (ed) *Essays in crop plant evolution*. Cambridge University Press, Cambridge, pp 123–141
- Doebley J (2006) Unfallen grains: how ancient farmers turned weeds into crops. *Science* 312:1318–1319
- Doebley J, Stec A, Hubbard L (1997) The evolution of apical dominance in maize. *Nature* 386:485–488. doi:[10.1038/386485a0](https://doi.org/10.1038/386485a0)
- Doğanlar S, Frary A, Daunay M-C, Huvenaars K, Mank R, Frary A (2014) High resolution map of eggplant (*Solanum melongena*) reveals extensive chromosome rearrangement in domesticated members of the Solanaceae. *Euphytica* 198:231–241
- Douche DS, Graefius EJ (2005) Transformation for insect resistance. In: Razdan MK, Mattoo AK (eds) *Genetic improvement of Solanaceous crops*, vol 1, Potato. Science, Enfield, pp 235–266
- Dounce AL (1952) Duplicating mechanism for peptide chain and nucleic acid synthesis. *Enzymologia* 15:251–258
- Druka A, Potokina E, Luo Z, Jiang N, Chen X, Kearsey M, Waugh R (2010) Expression quantitative trait loci analysis in plants. *Plant Biotechnol J* 8:10–27
- Du J, Verzaux E, Chaparro-Garcia A, Bijsterbosch G, Keizer LCP, Zhou J, Liebrand TWH, Xie C, Govers F, Robatzek S, van der Vossen EAG, Jacobsen E, Visser RGF, Kamoun S, Vleeshouwers VGAA (2015) Elicitin recognition confers enhanced resistance to *Phytophthora infestans* in potato. *Nature Plants* 1: Article number: 15034. doi:[10.1038/nplants.2015.34](https://doi.org/10.1038/nplants.2015.34)
- Dudley JW (2007) From means to QTL: the Illinois long-term selection experiment as a case study in quantitative genetics. *Crop Sci* 47(Suppl 3):S20–S31. doi:[10.2135/cropsci2007.04.0003IPBS](https://doi.org/10.2135/cropsci2007.04.0003IPBS)
- Dudley JW (2008) Epistatic interactions in crosses of Illinois high oil × Illinois low oil and of Illinois high protein × Illinois low protein corn strains. *Crop Sci* 48:59–68. doi:[10.2135/cropsci2007.04.0242](https://doi.org/10.2135/cropsci2007.04.0242)
- Dudley JW, Lambert RJ (2004) 100 generations of selection for oil and protein in corn. *Plant Breed Rev* 24:79–110
- Durel CE, Laurens F, Fouillet A, Lespinasse Y (1998) Utilization of pedigree information to estimate genetic parameters from large unbalanced data sets in apple. *Theor Appl Genet* 96:1077–1085
- East EM (1908) Inbreeding in corn. In Connecticut Agricultural Experiments Station Report 1907, pp 419–428

- East EM (1915) Studies on size inheritance in *Nicotiana*. *Genetics* 1:164–176
- Eberhart SA, Russell WA, Penny LH (1964) Double cross hybrid prediction in maize when epistasis is present. *Crop Sci* 4:363–366
- EFSA (2010) EFSA Panel on Genetically Modified Organisms (GMO). Guidance on the environmental risk assessment of genetically modified plants. *EFSA J* 8(11): 1879
- EFSA (2011) EFSA Panel on Genetically Modified Organisms (GMO). Guidance for risk assessment of food and feed from genetically modified plants. *EFSA J* 9(5): 2150
- EFSA (2012a) EFSA Panel on Genetically Modified Organisms (GMO). Scientific opinion addressing the safety assessment of plants developed through cisgenesis and intragenesis. *EFSA J* 10(2):2561
- EFSA (2012b) EFSA Panel on Genetically Modified Organisms (GMO). Scientific opinion addressing the safety assessment of plants developed using Zinc Finger Nuclease 3 and other Site-Directed Nucleases with similar function. *EFSA J* 10(10):2943
- Eichten SR, Swanson-Wagner RA, Schnable JC, Waters AJ, Hermanson PJ, Liu S, Yeh C-J, Gendler K, Freeling M, Schnable PS, Vaughn MW, Springer NM (2011) Heritable epigenetic variation among maize inbreds. *PLoS Genet* 7, e1002372
- Ellerström S, Zagorcheva L (1977) Sterility and apomictic embryo-sac formation in *Raphanobrassica*. *Hereditas* 87:107–119
- Ellis RT (1995) Tea. In: Smartt J, Simmonds NW (eds) *Evolution of crop plants*, 2nd edn. Longman Scientific & Technical, Harlow, pp 22–27
- Ellis JG, Lawrence GJ, Luck JE, Dodds PN (1999) Identification of regions in alleles of the flax rust resistance gene *L* that determine differences in gene-for-gene specificity. *Plant Cell* 11:495–506
- Emerson RA (1911) Genetic correlation and spurious allelomorphism in maize. Annual Report of the Nebraska Agricultural Experiment Station 24: 58–90
- Esquinás-Alcázar JT, Frison C, López F (2011) Introduction. In: Frison C, López F, Esquinás-Alcázar JT (eds) *Plant genetic resources and food security*. FAO, Biodiversity International and Earthscan, Abingdon, pp 1–23
- Evans LT (1996) *Crop evolution, adaptation and yield*. Cambridge University Press, Cambridge, 500p
- Eversole K (2012) The combined strategies of the IWGSC to achieve a reference sequence of the bread wheat genome. In: Bedő Z, Láng L (eds) *Plant breeding for future generations*. Proceedings of the 19th EUCARPIA General Congress. Agricultural Institute, Centre for Agricultural Research, Hungarian Academy of Sciences, Martonvásár, pp 58–61
- Ewen SWB, Pusztai A (1999) Effect of diet containing genetically modified potatoes expressing *Galanthus nivalis* lectin on rat small intestine. *Lancet* 354:1353–1354
- Fahim M, Dhanapala MP, Senadhira D, Lawrence MJ (1998) Quantitative genetics of rice II. A comparison of the efficiency of four breeding methods. *Field Crop Res* 55:257–266
- Falconer DS, Mackay TFC (1996) *Introduction to quantitative genetics*, 4th edn. Longman, Harlow, 464p
- Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164:1567–1587
- FAO (2013) Genebank standards for plant genetic resources for food and agriculture. FAO, Rome, 166p
- FAO, IFAD, WFP (2015) The state of food insecurity in the world 2015—Meeting the 2015 international hunger targets: taking stock of uneven progress. FAO, Rome, 56p
- Fehr WR, Hadley HH (eds) (1980) *Hybridization of crop plants*. American Society of Agronomy and Crop Science Society of America, Madison, 765p
- Fernández-Martínez JM, Pérez-Vich B, Velasco L (2009) Sunflower. In: Vollmann J, Rajcan I (eds) *Oil crops*, vol 4, *Handbook of plant breeding*. Springer, New York, pp 155–232
- Finlay KW, Wilkinson GN (1963) The analysis of adaptation in a plant-breeding programme. *Aust J Agric Res* 14:742–754

- Finn CE, Hancock JF (2008) Raspberries. In: Hancock JF (ed) Temperate fruit crop breeding. Springer, New York, pp 359–392
- Finney DJ (1958) Plant selection for yield improvement. *Euphytica* 7:83–106
- Fisher RA (1918) The correlation between relatives on the supposition of Mendelian inheritance. *Trans R Soc Edinb* 52:399–433
- Fisher RA, Mather K (1943) The inheritance of style length in *Lythrum salicaria*. *Ann Eugen* 12:1–23
- Fisher RA, Yates F (1963) Statistical tables, 6th edn. Oliver & Boyd, Edinburgh, 146p
- Fitch MMM, Manshardt RM, Gonsalves D, Slightom JL, Sanford JC (1992) Virus resistant papaya derived from tissues bombarded with the coat protein gene of papaya ringspot virus. *Nat Biotechnol* 10:1466–1472
- Flor HH (1942) Inheritance of pathogenicity in *Melampsora lini*. *Phytopathology* 32:653–659
- Flor HH, Comstock VE (1971) Development of flax cultivars with multiple rust-conditioning genes. *Crop Sci* 11:64–66
- Fonceka D, Tossim HA, Rivallan R, Vignes H, Lacut E, de Bellis F, Faye I, Ndoye O, Leal-Bertioli SCM, Valls JFM, Bertioli DJ, Glaszmann JC, Courtois B, Rami JF (2012) Construction of chromosome segment substitution lines in peanut (*Arachis hypogaea* L.) using a wild synthetic and QTL mapping for plant morphology. *PLoS One* 7(11), e48642. doi:[10.1371/journal.pone.0048642](https://doi.org/10.1371/journal.pone.0048642)
- Forster BP, Franckowiak JD, Lundqvist U, Thomas WTB, Leader D, Shaw P, Lyon J, Waugh R (2012) Mutant phenotyping and pre-breeding in barley. In: Shu QY, Forster BP, Nakagawa H (eds) Plant mutation breeding and biotechnology. CABI, Wallingford, pp 327–346
- Frascaroli E, Cane MA, Landi P, Pea G, Gianfranceschi L, Villa M, Morgante M, Pe ME (2007) Classical genetic and quantitative trait loci analyses of heterosis in a maize hybrid between two elite inbred lines. *Genetics* 176:625–644
- Freialdenhoven A, Peterhansel C, Kurth J, Kreuzaler F, Schulze-Lefert P (1996) Identification of genes required for the function of non-race-specific Mlo resistance to powdery mildew in barley. *Plant Cell* 8:5–14
- Friebe B, Jiang J, Raupp WJ, McIntosh RA, Gill BS (1996) Characterization of wheat-alien translocations conferring resistance to diseases and pests: current status. *Euphytica* 91:59–87
- Friedberg EC, Walker GC, Siede W, Wood RD, Schultz RA, Ellenberger T (2006) DNA repair and mutagenesis, 2nd edn. ASM, Washington, DC, 1118p
- Friedt W, Snowdon R (2009) Oilseed rape. In: Vollmann J, Rajcan I (eds) Oil crops, vol 4, Handbook of plant breeding. Springer, New York, pp 91–126
- Frison C, López F, Esquinas-Alcázar JT (eds) (2011) Plant genetic resources and food security. FAO, Bioversity International and Earthscan, Abingdon, 321p
- Fritzsche-Neto R, Borém A (eds) (2012) Plant breeding for abiotic stress tolerance. Springer, Heidelberg, 175p
- Fu TD, Yang GS, Yang XN (1990) Studies on three line Polima cytoplasmic male sterility developed in *B. napus*. *Plant Breed* 104:115–120
- Fu H-W, Li Y-F, Shu Q-Y (2008) A revisit of mutation induction by gamma rays in rice (*Oryza sativa* L.): implications of microsatellite markers for quality control. *Mol Breed* 22:281–288. doi:[10.1007/s11032-008-9173-7](https://doi.org/10.1007/s11032-008-9173-7)
- Furuta T, Uehara K, Angeles-Shim RB, Shim J, Ashikari M, Takashi T (2014) Development and evaluation of chromosome segment substitution lines (CSSLs) carrying chromosome segments derived from *Oryza rufipogon* in the genetic background of *Oryza sativa* L. *Breed Sci* 63:468–475. doi:[10.1270/jsbbs.63.468](https://doi.org/10.1270/jsbbs.63.468)
- Gale JS (1980) Population genetics. Blackie, Glasgow, 321p
- Gale JS (1990) Theoretical population genetics. Unwin Hyman, London, 417p
- Gallais A (2003) Quantitative genetics and breeding methods in autopolyploid plants. INRA, Paris, 515p
- Gamow G (1954) Possible relation between deoxyribonucleic acid and protein structures. *Nature* 173:318

- Gamuyao R, Chin JH, Pariasca-Tanaka J, Pesaresi P, Catausan S, Dalid C, Slamet-Loedin I, Tecson-Mendoza EM, Wissuwa M, Heuer S (2012) The protein kinase Pstol1 from traditional rice confers tolerance of phosphorus deficiency. *Nature* 488:535–539. doi:[10.1038/nature11346](https://doi.org/10.1038/nature11346)
- Garavito A, Guyot R, Lozano J, Gavory F, Samain S, Panaud O, Tohme J, Ghesquière A, Lorieux M (2010) A genetic model for the female sterility barrier between Asian and African cultivated rice species. *Genetics* 185:1425–1440
- Garcia AAF, Wang S, Melchinger AE, Zeng ZB (2008) Quantitative trait loci mapping and the genetic basis of heterosis in maize and rice. *Genetics* 180:1707–1724
- Gassmann AJ, Petzold-Maxwell JL, Clifton EH, Dunbar MW, Hoffmann AM, Ingber DA, Keweshan RS (2014) Field-evolved resistance by western corn rootworm to multiple *Bacillus thuringiensis* toxins in transgenic maize. *Proc Natl Acad Sci* 111:5141–5146
- Gauch HG Jr (1988) Model selection and validation for yield trials with interaction. *Biometrics* 44:705–715
- Gauch HG Jr, Piepho HP, Annicchiarico P (2008) Statistical analysis of yield trials by AMMI and GGE: further considerations. *Crop Sci* 48:866–889
- Gaul H (1960) Critical analysis of the methods for determining the mutation frequency after seed treatment with mutagens. *Genet Agrar* 12:297–318
- Gaul H (1964) Mutations in plant breeding. *Radiat Bot* 4:155–232
- Gaur PC, Pandey SK (2000) Potato improvement in sub-tropics. In: Khurana SMP, Shekhawat GS, Singh BP et al (eds) Potato, global research & development, vol 1. Indian Potato Association, Shimla, pp 52–63
- Gavrilenko T (2011) Application of molecular cytogenetics in fundamental and applied research of potato. In: Bradeen JM, Kole C (eds) Genetics, genomics and breeding of potato. Science, Enfield, pp 184–206
- Geiger HH, Miedaner T (2009) Rye breeding. In: Carena MJ (ed) Cereals, vol 3, Handbook of plant breeding. Springer, New York, pp 157–181
- Ghesquière M, Humphreys MW, Zwierzykowski Z (2010) *Festulolium*. In: Boller B, Possekt UK, Veronesi F (eds) Fodder crops and amenity grasses, vol 5, Handbook of plant breeding. Springer, New York, pp 293–316
- Ghislain M, Montenegro JD, Juarez H, del Rosario HM (2014) Ex-post analysis of landraces sympatric to a commercial variety in the center of origin of the potato failed to detect gene flow. *Transgenic Res.* doi:[10.1007/s11248-014-9854-4](https://doi.org/10.1007/s11248-014-9854-4)
- Gilbert N (2009) The disappearing nutrient. *Nature* 461:716–718
- Gilbert N (2013) Case studies: a hard look at GM crops. *Nature* 497:24–26
- Gilbert N (2014) Cross-bred crops get fit faster: genetic engineering lags behind conventional breeding in efforts to create drought-resistant maize. *Nature* 513:292. doi:[10.1038/513292a](https://doi.org/10.1038/513292a)
- Gleeson AC (1997) Spatial analysis. In: Kempton RA, Fox PN (eds) Statistical methods for plant variety evaluation. Chapman & Hall, London, pp 68–85
- Glendinning DR (1976) Neo-Tuberousum: new potato breeding material. 4. The breeding system of Neo-Tuberousum, and the structure and composition of the Neo-Tuberousum gene pool. *Potato Res* 19:27–36
- Glendinning DR (1983) Potato introductions and breeding up to the early 20th century. *New Phytol* 94:479–505
- Godfray HCJ (2014) How can 9–10 billion people be fed sustainably and equitably by 2050? In: Goldin I (ed) Is the planet full? Oxford University Press, Oxford, pp 104–120
- Goldstein WA, Schmidt W, Burger H, Messmer M, Pollak LM, Smith ME, Goodman MM, Kutka FJ, Pratt RC (2012) Maize: breeding and field testing for organic farmers. In: Lammerts van Bueren ET, Myers JR (eds) Organic crop breeding. Wiley, Chichester, pp 175–189
- Goodwin SB, Drenth A (1997) Origin of the A2 mating type of *Phytophthora infestans* outside Mexico. *Phytopathology* 87:992–999

- Goulden CH (1939) Problems in plant selections. In: Punnett RC (ed) Proceedings of the 7th international genetical congress, Edinburgh. Cambridge University Press, Cambridge, pp 132–133
- Goverse A, Struik PC (2009) Debate on the exploitation of natural plant diversity to create late blight resistance in potato. *Potato Res* 52:265–271
- Gowers S (1981) Self-pollination in swedes (*Brassica napus* ssp. *rapifera*) and its implications for cultivar production. *Euphytica* 30:813–817
- Grant V (1971) Plant speciation. Columbia University Press, New York, 435p
- Green RE, Cornell SJ, Scharlemann JPW, Balmford A (2005) Farming and the fate of wild nature. *Science* 307:550–555
- Greene EA, Codomo CA, Taylor NE, Henikoff JG, Till BJ, Reynolds SH, Enns LC, Burtner C, Johnson JE, Odden AR, Comai L, Henikoff S (2003) Spectrum of chemically induced mutations from a large-scale reverse-genetic screen in *Arabidopsis*. *Genetics* 164:731–740
- Griffing B (1956) Concept of general and specific combining ability in relation to diallel crossing systems. *Aust J Biol Sci* 9:463–493
- Guha S, Maheshwari SC (1964) *In vitro* production of embryos from anthers of *Datura*. *Nature* 204:497
- Guimaraes EP (2012) Rice breeding. In: Carena MJ (ed) Cereals, vol 3, Handbook of Plant Breeding. Springer, New York, pp 99–126
- Gullì M, Salvatori E, Fusaro L, Pellacani C, Manes F, Marmiroli N (2015) Comparison of drought stress response and gene expression between a GM maize variety and a near-isogenic non-GM variety. *PLoS One* 10(2), e0117073. doi:[10.1371/journal.pone.0117073](https://doi.org/10.1371/journal.pone.0117073)
- Guo ZA, Song YX, Zhou RH, Ren ZL, Jia JZ (2010) Discovery, evaluation and distribution of haplotypes of the wheat *Ppd-D1* gene. *New Phytol* 185:841–851
- Gupta PK, Rustgi S, Mir RR (2008) Array-based high-throughput DNA markers for crop improvement. *Heredity* 101:5–18. doi:[10.1038/hdy.2008.35](https://doi.org/10.1038/hdy.2008.35)
- Gustafsson Å, MacKey J (1948) The genetical effects of mustard gas substances and neutrons. *Hereditas* 34:371–386
- Hackett CA, Bradshaw JE, Meyer RC, McNicol JW, Milbourne D, Waugh R (1998) Linkage analysis in tetraploid species: a simulation study. *Genet Res* 71:143–154
- Hackett CA, Bradshaw JE, McNicol JW (2001) Interval mapping of quantitative trait loci in autotetraploid species. *Genetics* 159:1819–1832
- Hackett CA, Milne I, Bradshaw JE, Luo Z (2007) TetraploidMap for windows: linkage map construction and QTL mapping in autotetraploid species. *J Hered* 98:727–729
- Hackett CA, McLean K, Bryan GJ (2013) Linkage analysis and QTL mapping using SNP dosage data in a tetraploid potato mapping population. *PLoS One* 8(5), e63939. doi:[10.1371/journal.pone.0063939](https://doi.org/10.1371/journal.pone.0063939)
- Hackett CA, Bradshaw JE, Bryan GJ (2014) QTL mapping in autotetraploids using SNP dosage information. *Theor Appl Genet* 127:1885–1904. doi:[10.1007/s00122-014-2347-2](https://doi.org/10.1007/s00122-014-2347-2), Epub 2014 Jul 1
- Hahlbrock K (2009) Feeding the planet: environmental protection through sustainable agriculture. Haus, London, 266p
- Hajjar R, Hodgkin T (2007) The use of wild relatives in crop improvement: a survey of developments over the last 20 years. *Euphytica* 156:1–13
- Haldane JBS (1919) The probable errors of calculated linkage values, and the most accurate method of determining gametic from certain zygotic series. *J Genet* 8:291–297
- Haldane JBS (1930) Theoretical genetics of autotetraploids. *J Genet* 22:359–372
- Haldane JBS (1931) The cytological basis of genetical interference. *Cytologia* 3:54–65
- Hallauer AR, Carena MJ (2009) Maize breeding. In: Carena MJ (ed) Cereals, vol 3, Handbook of plant breeding. Springer, New York, pp 3–98
- Hallauer AR, Eberhart SA (1970) Reciprocal full-sib selection. *Crop Sci* 10:315–316
- Hallauer AR, Carena MJ, Miranda FJB (2010) Quantitative genetics in maize breeding, 3rd edn, vol 6, Handbook of plant breeding. Springer, New York, 500p

- Hancock JF (2012) Plant evolution and the origin of crop species, 3rd edn. CABI, Wallingford, 245p
- Hancock JF, Luby JJ, Brown SK, Lobos GA (2008a) Apples. In: Hancock JF (ed) Temperate fruit crop breeding. Springer, New York, pp 1–37
- Hancock JF, Sjulin TM, Lobos GA (2008b) Strawberries. In: Hancock JF (ed) Temperate fruit crop breeding. Springer, New York, pp 393–437
- Hansen BG, Halkier BA, Kliebenstein DJ (2008) Identifying the molecular basis of QTLs: eQTLs add a new dimension. *Trends Plant Sci* 13:72–77
- Hardon JJ (1995) Oil palm. In: Smartt J, Simmonds NW (eds) Evolution of crop plants, 2nd edn. Longman Scientific & Technical, Harlow, pp 395–398
- Harlan JR (1992) Crops and man. American Society of Agronomy, Madison
- Harlan JR, de Wet JMJ (1971) Toward a rational classification of cultivated plants. *Taxon* 20:509–517
- Harlan HV, Martini ML (1929) A composite hybrid mixture. *Agron J* 21:487–490
- Harrington J (2012) Pioneer Hi-Bred, a DuPont business—bringing seed value to the grower. In: Acquaah G (ed) Principles of plant genetics and breeding. Wiley-Blackwell, Chichester, pp 357–360
- Harter AV, Gardner KA, Falush D, Lentz DL, Bye RA, Rieseberg LH (2004) Origin of extant domesticated sunflowers in eastern North America. *Nature* 430:201–205
- Hartl DL, Clark AG (1997) Principles of population genetics, 3rd edn. Sinauer, Sunderland, 542p
- Hawkes JG (1990) The potato: evolution, biodiversity & genetic resources. Belhaven, London, 259p
- Hawkes JG (1991) The importance of genetic resources in plant breeding. *Biol J Linn Soc* 43:3–10
- Hawkes JG, Francisco-Ortega J (1992) The potato in Spain during the late 16th century. *Econ Bot* 46:86–97
- Hawkes JG, Francisco-Ortega J (1993) The early history of the potato in Europe. *Euphytica* 70:1–7
- Hawkes JG, Jackson MT (1992) Taxonomic and evolutionary implications of the endosperm balance number hypothesis in potatoes. *Theor Appl Genet* 84:180–185
- Hawtin G, Fowler L (2011) The global crop diversity trust. In: Frison C, López F, Esquinas-Alcázar JT (eds) Plant genetic resources and food security. FAO, Biodiversity International and Earthscan, Abingdon, pp 209–221
- Hazel LN, Lush JL (1942) The efficiency of three methods of selection. *J Hered* 33:393–399
- Hedrick PW (2011) Genetics of populations, 4th edn. Jones and Bartlett, Sudbury, 675p
- Heldt H-W, Piechulla B (2011) Plant biochemistry, 4th edn. Academic, London, 622p
- Helgeson JP, Pohlman JD, Austin S, Haberlach GT, Wielgus SM, Ronis D, Zambolim L, Tooley P, McGrath JM, James RV, Stevenson WR (1998) Somatic hybrids between *Solanum bulbocastanum* and potato: a new source of resistance to late blight. *Theor Appl Genet* 96:738–742
- Hellens RP, Moreau C, Lin-Wang K, Schwinn KE, Thomson SJ, Fiers MWEJ, Frew TJ, Murray SR, Hofer JMI, Jacobs JME, Davies KM, Allan AC, Bendahmane A, Coyne CJ, Timmerman-Vaughan GM, Ellis THN (2010) Identification of Mendel's white flower character. *PLoS One* 5 (10), e13230. doi:[10.1371/journal.pone.0013230](https://doi.org/10.1371/journal.pone.0013230)
- Henderson CR (1950) Estimation of genetic parameters. *Ann Math Stat* 21:309–310
- Henry M, Beguin M, Requier F, Rollin O, Odoux J-F, Aupinel P, Aptel J, Tchamitchian S, Decourtey A (2012) A common pesticide decreases foraging success and survival in honey bees. *Science* 336:348–350. doi:[10.1126/science.1215039](https://doi.org/10.1126/science.1215039)
- Hermsen JGT (1994) Introgression of genes from wild species, including molecular and cellular approaches. In: Bradshaw JE, Mackay GR (eds) Potato genetics. CAB International, Wallingford, pp 515–538
- Hickey M, King C (2000) The Cambridge illustrated glossary of botanical terms. Cambridge University Press, Cambridge, 208p
- Hijmans RJ, Spooner DM, Salas AR, Guarino L, de la Cruz J (2002) Atlas of wild potatoes. Systematic and ecogeographic studies in crop genepools 8. IPGRI, Rome

- Hijmans RJ, Gavrilenko T, Stephenson S, Bamberg J, Salas A, Spooner DM (2007) Geographical and environmental range expansion through polyploidy in wild potatoes (*Solanum* section *Petota*). *Glob Ecol Biogeogr* 16:485–495
- Hillman G, Hedges R, Moore A, Colledge S, Pettitt P (2001) New evidence of Lateglacial cereal cultivation at Abu Hureyra on the Euphrates. *The Holocene* 11:383–393
- Hirsch CN, Foerster JM, Johnson JM, Sekhon RS, Muttoni G, Vaillancourt B, Peñagaricano F, Lindquist E, Pedraza MA, Barry K, de Leon N, Kaepler SM, Buella CR (2014) Insights into the maize pan-genome and pan-transcriptome. *Plant Cell* 26(1):121–135. doi:[10.1105/tpc.113.119982](https://doi.org/10.1105/tpc.113.119982)
- Hobhouse H (1992) Seeds of change. Papermac, London, 252p
- Hoerl AE, Kennard RW (1970) Ridge regression: biased estimation for nonorthogonal problems. *Technometrics* 12:55–67
- Holton TA, Cornish EC (1995) Genetics and biochemistry of anthocyanin biosynthesis. *Plant Cell* 7:1071–1083
- Horneburg B, Myers JR (2012) Tomato: breeding for improved disease resistance in fresh market and home garden varieties. In: Lammerts van Bueren ET, Myers JR (eds) Organic crop breeding. Wiley, Chichester, pp 239–249
- Horsch RB, Fraley RT, Rogers SG, Sanders PR, Lloyd A, Hoffmann N (1984) Inheritance of functional foreign genes in plants. *Science* 223:496–498
- Horsley RD, Franckowiak JD, Schwarz PB (2009) Barley. In: Carena MJ (ed) Cereals, vol 3, Handbook of plant breeding. Springer, New York, pp 227–250
- Hospital F (2003) Marker-assisted breeding. In: Newbury HJ (ed) Plant molecular breeding. Blackwell, Oxford, pp 30–59
- Hougas RW, Peloquin SJ, Ross RW (1958) Haploids of the common potato. *J Hered* 49:103–107
- Howard A, Pelc SR (1951) Nuclear incorporation of P<sup>32</sup> as demonstrated by autoradiographs. *Exp Cell Res* 2:178–187
- Howden NJK, Burt TP, Worrall F, Mathias S, Whelan MJ (2011) Nitrate pollution in intensively farmed regions: what are the prospects for sustaining high-quality groundwater? *Water Resour Res* 47:W00L02. doi:[10.1029/2011WR010843](https://doi.org/10.1029/2011WR010843)
- Hoyt E (1988) Conserving the wild relatives of crops. IBPGR, Rome and IUCN/WWF, Gland
- Huaman Z, Golmirzaie A, Amoros W (1997) The potato. In: Fuccillo D, Sears L, Stapleton P (eds) Biodiversity in trust: conservation and use of plant genetic resources in CGIAR centres. Cambridge University Press, Cambridge, pp 21–28
- Huaman Z, Ortiz R, Gomez R (2000a) Selecting a *Solanum tuberosum* subsp. *andigena* core collection using morphological, geographical, disease and pest descriptors. *Am J Potato Res* 77:183–190
- Huaman Z, Ortiz R, Zhang D, Rodriguez F (2000b) Isozyme analysis of entire and core collections of *Solanum tuberosum* subsp. *andigena* potato cultivars. *Crop Sci* 40:273–276
- Huang S, Sirikhachornkit A, Su X, Faris J, Gill B, Haselkorn R, Gornicki P (2002) Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the Triticum/Aegilops complex and the evolutionary history of polyploid wheat. *Proc Natl Acad Sci* 99:8133–8138
- Huang P, Molina J, Flowers JM, Rubinstein S, Jackson SA, Purugganan MD, Schaal BA (2012a) Phylogeography of Asian wild rice, *Oryza rufipogon*: a genome-wide view. *Mol Ecol* 21:4593–4604. doi:[10.1111/j.1365-294X.2012.05625.x](https://doi.org/10.1111/j.1365-294X.2012.05625.x)
- Huang X, Kurata N, Wei X, Wang ZX, Wang A, Zhao Q, Zhao Y, Liu K, Lu H, Li W, Guo Y, Lu Y, Zhou C, Fan D, Weng Q, Zhu C, Huang T, Zhang L, Wang Y, Feng L, Furuumi H, Kubo T, Miyabayashi T, Yuan X, Xu Q, Dong G, Zhan Q, Li C, Fujiyama A, Toyoda A, Lu T, Feng Q, Qian Q, Li J, Han B (2012b) A map of rice genome variation reveals the origin of cultivated rice. *Nature* 490:497–501
- Huff DR (2010) Bluegrasses. In: Boller B, Posselt UK, Veronesi F (eds) Fodder crops and amenity grasses, vol 5, Handbook of plant breeding. Springer, New York, pp 345–379
- Huffman MB, Xu X, Van Heerwaarden J, Pyhäjärvi T, Chia J-M, Cartwright RA, Elshire RJ, Glaubitz JC, Guill KE, Kaepler SM, Lai J, Morrell PL, Shannon LM, Song C, Springer NM,

- Swanson-Wagner RA, Tiffin P, Wang J, Zhang G, Doebley J, McMullen MD, Ware D, Buckler ES, Yang S, Ross-Ibarra J (2012) Comparative population genomics of maize domestication and improvement. *Nat Genet* 44:808–811
- Humphreys M, Feuerstein U, Vandewalle M, Baert J (2010) Ryegrasses. In: Boller B, Posselt UK, Veronesi F (eds) *Fodder crops and amenity grasses*, vol 5, *Handbook of plant breeding*. Springer, New York, pp 211–260
- Hung H-Y, Shannon LM, Tian F, Bradbury PJ, Chen C, Flint-Garcia SA, McMulleng MD, Ware D, Buckler ES, Doebley JF, Holland JB (2012a) ZmCCT and the genetic basis of day-length adaptation underlying the postdomestication spread of maize. *Proc Natl Acad Sci* 109:E1913–E1921
- Hung H-Y, Browne C, Guill K, Coles N, Eller M, Garcia A, Lepak N, Melia-Hancock S, Oropenza-Rosas M, Salvo S, Upadyayula N, Buckler ES, Flint-Garcia S, McMullen MD, Rocheford TR, Holland JB (2012b) The relationship between parental genetic or phenotypic divergence and progeny variation in the maize nested association mapping population. *Heredity* 108:490–499
- Hunter D, Heywood V (eds) (2011) *Crop wild species: a manual of *in situ* conservation*. Earthscan, London, 414p
- Ibarra-Perez FJ, Ehdaie B, Waines JG (1997) Estimation of outcrossing rate in common bean. *Crop Sci* 37:60–65
- Ishii T, Yonezawa K (2006) Optimizing the procedure of multi-stage yield selection for selecting the best from a population of candidate entries. *Breed Sci* 56:7–15
- Jaccoud D, Peng K, Feinstein D, Kilian A (2001) Diversity arrays: a solid state technology for sequence information independent genotyping. *Nucleic Acids Res* 29, e25
- Jackson RC (1957) New low chromosome number for plants. *Science* 126:1115–1116
- Jacobsen E (2007) The canon of potato science. 6. Genetic modification and cis- and transgenesis. *Potato Res* 50:227–230
- Jaggard KW, Qi A, Ober ES (2010) Possible changes to arable crop yields by 2050. *Phil Trans R Soc B* 365:2835–2851
- Jagosz B (2011) The relationship between heterosis and genetic distances based on RAPD and AFLP markers in carrot. *Plant Breed* 130:574–579. doi:[10.1111/j.1439-0523.2011.01877.x](https://doi.org/10.1111/j.1439-0523.2011.01877.x)
- James C (2014) Global status of commercialized Biotech/GM Crops: 2014. ISAAA Brief No. 49. ISAAA, Ithaca, 259p
- Jan CC, Vick BA (2006) Registration of seven cytoplasmic male-sterile and four fertility restoration sunflower germplasms. *Crop Sci* 46:1829–1830
- Janick J (ed) (2004) *Plant breeding reviews*, vol 24, Part 1: Long-term selection: maize. Wiley, Hoboken, 361p
- Jansen RC (1993) Interval mapping of multiple quantitative trait loci. *Genetics* 135:205–211
- Jansky S (2006) Overcoming hybridization barriers in potato. *Plant Breed* 125:1–12
- Jansky S (2009) Breeding, genetics and cultivar development. In: Singh J, Kaur L (eds) *Advances in potato chemistry and technology*. Academic, Burlington, pp 27–62
- Jansky SH, Davis GL, Peloquin SJ (2004) A genetic model for tuberization in potato haploid-wild species hybrids grown under long-day conditions. *Am J Potato Res* 81:335–339
- Janssens FA (1909) Spermatogénèse dans les Batraciens. V La théorie de la chiasmatypie Nouvelles interprétation des cinèses de maturation. *Cellule* 25:387–411
- Jarvis DI, Myer L, Klemick H, Guarino L, Smale M, Brown AHD, Sadiki M, Sthapit B, Hodgkin T (2000) A Training guide for *in situ* conservation on-farm. Version 1. International Plant Genetic Resources Institute, Rome
- Jarvis A, Lane A, Hijmans RJ (2008) The effect of climate change on crop wild relatives. *Agric Ecosyst Environ* 126:13–23
- Jeffries C, Barker H, Khurana SMP (2006) Viruses and viroids. In: Gopal J, Khurana SMP (eds) *Handbook of potato production, improvement, and postharvest management*. Food Products, New York, pp 387–448
- Jenkin TJ (1933) Interspecific and intergeneric hybrids in herbage grasses. Initial crosses. *J Genet* 28:205–264

- Jenkins MT (1934) Methods of estimating the performance of double crosses in corn. *J Am Soc Agron* 26:199–204
- Jensen NF (1988) Plant breeding methodology. Wiley, New York, pp 271–296
- Jensen NF, Federer WT (1964) Adjacent row competition in wheat. *Crop Sci* 4:641–645
- Jensen SD, Kuhn WE, McConnell RL (1983) Combining ability in elite U.S. maize germplasm. In: Proc. Annu. Corn Sorghum Ind. Res. Conf. 38: 87–96
- Jeukens M (2012) Breeding for durable resistance against an oomycete in lettuce. In: Acquaah G (ed) Principles of plant genetics and breeding. Wiley-Blackwell, Chichester, pp 273–276
- Jia Y (2009) Artificial introgression of a large chromosome fragment around the rice blast resistance gene *Pi-ta* in backcross progeny and several elite rice cultivars. *Heredity* 103:333–339. doi:10.1038/hdy.2009.95
- Jin LP, Qu DY, Xie KY, Bian CS, Duan SG (2004) Potato germplasm, breeding studies in China. In: Proceedings of the fifth world potato congress. Kunming, pp 175–178
- Jin L, Zhang H, Lu Y, Yang Y, Wu K, Tabashnik BE, Wu Y (2014) Large-scale test of the natural refuge strategy for delaying insect resistance to transgenic Bt crops. *Nat Biotechnol.* doi:10.1038/nbt.3100
- Jinks JL (1983) Biometrical genetics of heterosis. In: Frankel R (ed) Heterosis: reappraisal of theory and practice, vol 6, Monographs on theoretical and applied genetics., pp 1–46
- Jinks JL, Connolly V (1973) Selection for specific and general response to environmental differences. *Heredity* 30:33–40
- Jinks JL, Connolly V (1975) Determination of the environmental sensitivity of selection lines by the selection environment. *Heredity* 34:401–406
- Jinks JL, Pooni HS (1980) Comparing predictions of mean performance and environmental sensitivity of recombinant inbred lines based upon  $F_3$  and triple test cross families. *Heredity* 45:305–312
- Jinks JL, Pooni HS (1981a) Comparative results of selection in the early and late stages of an inbreeding programme. *Heredity* 46:1–7
- Jinks JL, Pooni HS (1981b) Properties of pure-breeding lines produced by dihaploidy, single seed descent and pedigree breeding. *Heredity* 46:391–395
- Jinks JL, Pooni HS (1982) Determination of the environmental sensitivity of selection lines of *Nicotiana rustica* by the selection environment. *Heredity* 49:291–294
- Jo K-R, Kim C-J, Kim S-J, Kim T-Y, Bergeroet M, Jongsma MA, Visser RGF, Jacobsen E, Vossen JH (2014) Development of late blight resistant potatoes by cisgene stacking. *BMC Biotechnol* 14: 50. <http://www.biomedcentral.com/1472-6750/14/50>
- Johannsen W (1909) Elemente der exakten Erblichkeitslehre. Fischer, Jena
- Johns T, Alonso JG (1990) Glycoalkaloid change during the domestication of the potato, *Solanum Section Petota*. *Euphytica* 50:203–210
- Johnson RC, Bradley VL, Evans MA (2002) Effective population size during grass germplasm seed regeneration. *Crop Sci* 42:286–290. doi:10.2135/cropsci2002.2860
- Johnston SA, der Nijls TPM, Peloquin SJ, Hanneman RE Jr (1980) The significance of genic balance to endosperm development in interspecific crosses. *Theor Appl Genet* 57:5–9
- Jonas E, de Koning D-J (2013) Does genomic selection have a future in plant breeding? *Trends Biotechnol* 31:497–504
- Jones DF (1917) Dominance of linked factors as a means of accounting for heterosis. *Genetics* 2:466–479
- Jones DF (1918) The effects of inbreeding and crossbreeding upon development. *Connecticut Agric Exp Stn Bull* 207:5–100
- Jones GH (1971) The analysis of exchanges in tritium-labelled meiotic chromosomes. II *Stethophyma grossum*. *Chromosoma* 34:367–382
- Jones JDG (2001) Putting knowledge of plant disease resistance genes to work. *Curr Opin Plant Biol* 4:281–287
- Jones JN, Clarke AE (1943) Inheritance of male sterility in the onion and the production of hybrid seed. *Proc Amer Soc Hort Sci* 43:189–194

- Jones LP, Compton WA, Gardner CO (1971) Comparison of full and half-sib reciprocal recurrent selection. *Theor Appl Genet* 41:36–39
- Jones MP, Dingkuhn M, Aluko GK, Semon M (1997) Interspecific *Oryza Sativa* L. X *O. Glaberrima* Steud progenies in upland rice improvement. *Euphytica* 92:237–246
- Jones H, Civáň P, Cockram J, Leigh FJ, Smith LMJ, Jones MK, Charles MP, Molina-Cano J-L, Powell W, Jones G, Brown TA (2011) Evolutionary history of barley cultivation in Europe revealed by genetic analysis of extant landraces. *BMC Evol Biol* 11:320. doi:[10.1186/1471-2148-11-320](https://doi.org/10.1186/1471-2148-11-320)
- Jones JDG, Witek K, Verweij W, Jupe F, Cooke D, Dorling S, Tomlinson L, Smoker M, Perkins S, Foster S (2014) Elevating crop disease resistance with cloned genes. *Philos Trans R Soc B* 369:20130087. doi:[10.1098/rstb.2013.0087](https://doi.org/10.1098/rstb.2013.0087)
- Jongedijk E, Ramanna MS (1988) Synaptic mutants in potato, *Solanum tuberosum* L. I. Expression and identity of genes for desynapsis. *Genome* 30:664–670
- Jongedijk E, Ramanna MS, Sawor Z, Hermsen JGT (1991) Formation of first division restitution (FDR) 2n-megaspores through pseudohomotypic division in *ds-1* (desynapsis) mutants of diploid potato: routine production of tetraploid progeny from 2x FDR × 2x FDR crosses. *Theor Appl Genet* 82:645–656
- Jorgensen JH (1992) Discovery, characterization and exploitation of *Mlo* powdery mildew resistance in barley. *Euphytica* 63:141–152
- Jumbo M, Weldekidan T, Holland JB, Hawk JA (2011) Comparison of conventional, modified single seed descent, and doubled haploid breeding methods for maize inbred line development using germplasm enhancement of maize breeding crosses. *Crop Sci* 51:1534–1543. doi:[10.2135/cropsci2010.10.0594](https://doi.org/10.2135/cropsci2010.10.0594)
- Jung CS, Griffiths HM, De Jong DM, Cheng S, Bodis M, De Jong WS (2005) The potato *P* locus codes for flavonoid 3',5' -hydroxylase. *Theor Appl Genet* 110:269–275
- Jung K-H, An G, Ronald PC (2008) Towards a better bowl of rice: assigning function to tens of thousands of rice genes. *Nat Rev Genet* 9:91–101
- Jung CS, Griffiths HM, De Jong DM, Cheng S, Bodis M, Kim TS, De Jong WS (2009) The potato *developer* (*D*) locus encodes an R2R3 MYB transcription factor that regulates expression of multiple anthocyanin structural genes in tuber skin. *Theor Appl Genet* 120:45–57
- Kaczmarczyk A, Rokka V-M, Keller ERJ (2011) Potato shoot tip cryopreservation. A review. *Potato Res* 54:45–79
- Kamiński P, Dyki B, Stępowska AA (2012) Improvement of cauliflower male sterile lines with *Brassica nigra* cytoplasm, phenotypic expression and possibility of practical application. *J Agric Sci* 4:190–200
- Kane NC, Burke JM, Marek L, Seiler G, Vear F, Baute G, Knapp SJ, Vincourt P, Rieseberg LH (2013) Sunflower genetic, genomic and ecological resources. *Mol Ecol Resour* 13:10–20
- Kang MS, Priyadarshan PM (eds) (2007) Breeding major food staples. Blackwell, Oxford, 437p
- Kavanagh E (2006) How and when was wild wheat domesticated? *Science* 313:296–297
- Kawano K, Amaya A, Daza P, Rios M (1978) Factors affecting efficiency of hybridization and selection in cassava. *Crop Sci* 18:373–376
- Kazama Y, Hirano T, Saito H, Liu Y, Ohbu S, Hayashi Y, Abe T (2011) Characterization of highly efficient heavy-ion mutagenesis in *Arabidopsis thaliana*. *BMC Plant Biol* 11:161. doi:[10.1186/1471-2229-11-161](https://doi.org/10.1186/1471-2229-11-161)
- Kearsey MJ (1985) The effect of linkage on additive genetic variance with inbreeding an *F<sub>2</sub>*. *Heredity* 55:139–143
- Kearsey MJ, Jinks JL (1968) A general method of detecting additive, dominance and epistatic variation for metrical traits: I Theory. *Heredity* 23:403–409
- Kearsey MJ, Pooni HS (1996) The genetical analysis of quantitative traits. Chapman & Hall, London, 381p
- Keightley PD (2004) Mutational variation and long-term selection response. *Plant Breed Rev* 24:227–247

- Kempler C, Hall H, Finn CE (2012) Raspberry. In: Badenes ML, Byrne DH (eds) *Fruit breeding*, vol 8, *Handbook of plant breeding*. Springer, New York, pp 263–304
- Kempthorne O (1957) An introduction to genetic statistics. Wiley, New York, 545p
- Kempton RA (1982) Adjustment for competition between varieties in plant breeding trials. *J Agric Sci (Camb)* 103:123–135
- Kempton RA (1984) The use of biplots in interpreting variety by environment interactions. *J Agric Sci (Camb)* 103:123–135
- Kempton RA (1997) Interference between plots. In: Kempton RA, Fox PN (eds) *Statistical methods for plant variety evaluation*. Chapman & Hall, London, pp 101–116
- Kempton RA, Lockwood G (1984) Interplot competition in variety trials of field beans (*Vicia faba* L.). *J Agric Sci (Camb)* 103:293–301
- Khachatourians GG, Sumner AK, Phillips PWB (2001) An Introduction to the history of canola and the scientific basis for innovation. In: Phillips PWB, Khachatourians GG (eds) *The biotechnology revolution in global agriculture: innovation and investment in the canola sector*. CAB International, Wallingford, pp 33–47
- Kiær LP, Skovgaard IM, Østergård H (2012) Effects of inter-varietal diversity, biotic stresses and environmental productivity on grain yield of spring barley variety mixtures. *Euphytica* 185:123–138. doi:[10.1007/s10681-012-0640-1](https://doi.org/10.1007/s10681-012-0640-1)
- King J (2011) Reaching for the sun. How plants work, 2nd edn. Cambridge University Press, Cambridge, 298p
- King IP, Morgan WG, Armstead IP, Harper JA, Hayward MD, Bolland A, Nash JV, Forster JW, Thomas HM (1998) Introgression mapping in the grasses. I Introgression of *Festuca pratensis* chromosomes and chromosome segments into *Lolium perenne*. *Heredity* 81:462–467. doi:[10.1046/j.1365-2540.1998.00437.x](https://doi.org/10.1046/j.1365-2540.1998.00437.x)
- King IP, Morgan WG, Harper JA, Thomas HM (1999) Introgression mapping in the grasses. II Meiotic analysis of the *Lolium perenne/Festuca pratensis* triploid hybrid. *Heredity* 82:107–112. doi:[10.1038/sj.hdy.6884680](https://doi.org/10.1038/sj.hdy.6884680)
- King J, Roberts LA, Kearsey MJ, Thomas HM, Jones RN, Huang L, Armstead IP, Morgan WG, King IP (2002) A demonstration of a 1:1 correspondence between chiasma frequency and recombination using a *Lolium perenne/Festuca pratensis* substitution. *Genetics* 161:307–314
- King J, Armstead IP, Donnison SI, Roberts LA, Harper JA, Skøt K, Elbrough K, King IP (2007) Comparative analyses between *Lolium/Festuca* introgression lines and rice reveal the major fraction of functionally annotated gene models is located in recombination-poor/very recombination-poor regions of the genome. *Genetics* 177:597–606
- Kingsbury N (2009) *Hybrid: the history & science of plant breeding*. The University of Chicago Press, Chicago, 493p
- Klein TM, Wolf ED, Wu R, Sanford JC (1987) High-velocity microprojectiles for delivering nucleic acids into living cells. *Nature* 327:70–73. doi:[10.1038/327070a0](https://doi.org/10.1038/327070a0)
- Kloosterman B, Abelenda JA, Carretero Gomez MdM, Oortwijn M, de Boer JM, Kowitzanich K, Horvath BM, van Eck HJ, Smacznak C, Prat S, Visser RGF, Bachem CWB (2013) Naturally occurring allele diversity allows potato cultivation in northern latitudes. *Nature* 495 (7440):246–250
- Klümpen W, Qaim M (2014) A meta-analysis of the impacts of genetically modified crops. *PLoS One* 9, e111629. doi:[10.1371/journal.pone.0111629](https://doi.org/10.1371/journal.pone.0111629)
- Knight TA (1807) On raising of new and early varieties of the potato (*Solanum tuberosum*). *Trans Hort Soc Lond* 1:57–59
- Knight R (1970) The measurement and interpretation of genotype-environment interactions. *Euphytica* 19:225–235
- Knight R (1973) The relation between hybrid vigour and genotype-environment interactions. *Theor Appl Genet* 43:311–318
- Kobayashi S, Goto-Yamamoto N, Hirochika H (2004) Retrotransposon-induced mutations in grape skin color. *Science* 304:982

- Kohli A, Leech M, Vain P, Laurie DA, Christou P (1998) Transgene organization in rice engineered through direct DNA transfer supports a two-phase integration mechanism mediated by the establishment of integration hot spots. *Proc Natl Acad Sci* 95:7203–7208
- Komatsuda T, Pourkheirandish M, He C, Azhagavel P, Kanamori H, Perovic D, Stein N, Graner A, Wicker T, Tagiri A, Lundqvist U, Fujimura T, Matsuoka M, Matsumoto T, Yano M (2007) Six-rowed barley originated from a mutation in a homeodomain-leucine zipper I-class homeobox gene. *Proc Natl Acad Sci* 104:1424–1429
- Konishi S, Izawa T, Lin SY, Ebana K, Fukuta Y, Sasaki T, Yano M (2006) An SNP caused loss of seed shattering during rice domestication. *Science* 312:1392–1396
- Kosambi DD (1944) The estimation of map distances from recombination values. *Ann Eugen* 12:172–175
- Kovach MJ, Sweeney MT, McCouch SR (2007) New insights into the history of rice domestication. *Trends Genet* 23:578–587
- Krishna V, Qaim M, Zilberman D (2015) Transgenic crops, production risk and agrobiodiversity. *Eur Rev Agric Econ.* doi:[10.1093/erae/jbv012](https://doi.org/10.1093/erae/jbv012). First published online: June 4, 2015
- Kumar S, Bink MCAM, Volz RK, Bus VGM, Chagné D (2012) Towards genomic selection in apple (*Malus × domestica* Borkh.) breeding programmes: Prospects, challenges and strategies. *Tree Genet Genomes* 8:1–14. doi:[10.1007/s11295-011-0425-z](https://doi.org/10.1007/s11295-011-0425-z)
- Kunter B, Bas M, Kantoglu Y, Burak M (2012) Mutation breeding of sweet cherry (*Prunus avium* L.) var. 0900 Ziraat. In: Shu QY, Forster BP, Nakagawa H (eds) Plant mutation breeding and biotechnology. CABI, Wallingford, pp 453–459
- Kyndt T, Quispe D, Zhai H, Jarret R, Ghislain M, Liu Q, Gheysen G, Kreuze JF (2015) The genome of cultivated sweet potato contains *Agrobacterium* T-DNA with expressed genes: an example of a naturally transgenic food crop. *Proc Natl Acad Sci* 112(18):5844–5849. doi:[10.1073/pnas.1419685112](https://doi.org/10.1073/pnas.1419685112)
- Lacombe S, Rougon-Cardoso A, Sherwood E, Peeters N, Dahlbeck D, van Esse HP, Smoker M, Rallapalli G, Thomma BPHJ, Staskawicz B, Jones JDG, Zipfel C (2010) Interfamily transfer of a plant pattern-recognition receptor confers broad-spectrum bacterial resistance. *Nat Biotechnol* 28:365–369. doi:[10.1038/nbt.1613](https://doi.org/10.1038/nbt.1613)
- Lai J, Li R, Xu X, Jin W, Xu M, Zhao H, Xiang Z, Song W, Ying K, Zhang M, Jiao Y, Ni P, Zhang J, Li D, Guo X, Ye K, Jian M, Wang B, Zheng H, Liang H, Zhang X, Wang S, Chen S, Li J, Fu Y, Springer NM, Yang H, Wang J, Dai J, Schnable PS, Wang J (2010) Genome-wide patterns of genetic variation among elite maize inbred lines. *Nat Genet* 42:1027–1030. doi:[10.1038/ng.684](https://doi.org/10.1038/ng.684)
- Landeo JA (2002) Durable resistance: quantitative/qualitative resistance. In: Lizarraga C (ed) Proceedings of the global initiative on late blight conference, Hamburg, 11–13 July 2002, pp 29–36
- Lander ES, Green P (1987) Construction of multilocus genetic linkage maps in humans. *Proc Natl Acad Sci* 84:2363–2367
- Lang J (2001) Notes of a potato watcher. Texas A&M University Press, College Station, 420p
- Latha R, Thiagarajan K, Senthivel S (2004) Genetics, fertility behaviour and molecular marker analysis of a new TGMS line, TS6, in rice. *Plant Breed* 123:235–240
- Lawrence MJ, Senadhira D (1998) Quantitative genetics of rice IV. A breeding strategy. *Field Crop Res* 55:275–281
- Lawrence MJ, Marshall DF, Davies P (1995) Genetics of genetic conservation. I. Sample size when collecting germplasm. *Euphytica* 84:89–99
- Lebot V (2010) Sweet potato. In: Bradshaw JE (ed) Root and tuber crops Handbook of plant breeding, vol 7. Springer, New York, pp 97–125
- Leclercq P (1969) Une sterilité mâle utilisable pour la production d'hybrides simples de tournesol. *Ann Amelior Plant* 16:135–144
- Lecomte L, Duffé P, Buret M, Servin B, Hospital F, Causse M (2004) Marker-assisted introgression of five QTLs controlling fruit quality traits into three tomato lines revealed interactions

- between QTLs and genetic backgrounds. *Theor Appl Genet* 109:658–668. doi:[10.1007/s00122-004-1674-0](https://doi.org/10.1007/s00122-004-1674-0)
- Lees AK, Stewart JA, Lynott JS, Carnegie SF, Campbell H, Roberts AMI (2012) The effect of a dominant *Phytophthora infestans* genotype (13\_A2) in Great Britain on host resistance to foliar late blight in commercial potato cultivars. *Potato Res* 55:125–134. doi:[10.1007/s11540-012-9214-9](https://doi.org/10.1007/s11540-012-9214-9)
- Leiser WL, Rattunde HFW, Piepho HP, Weltzien E, Diallo A, Melchinger AE, Parzies HK, Haussmann BIG (2012) Selection strategy for sorghum targeting phosphorus-limited environments in West Africa: analysis of multi-environment experiments. *Crop Sci* 52:2517–2527. doi:[10.2135/cropsci2012.02.0139](https://doi.org/10.2135/cropsci2012.02.0139)
- Leister D (2012) How can the light reactions of photosynthesis be improved in plants? *Front Plant Sci* 3:199. doi:[10.3389/fpls.2012.00199](https://doi.org/10.3389/fpls.2012.00199)
- León L, Velasco L, de la Rosa R (2015) Initial selection steps in olive breeding programs. *Euphytica* 201:453–462. doi:[10.1007/s10681-014-1232-z](https://doi.org/10.1007/s10681-014-1232-z)
- Lester DR, Ross JJ, Davies PJ, Reid JB (1997) Mendel's stem length gene (Le) encodes a gibberellin 3 beta-hydroxylase. *Plant Cell* 9:1435–1443
- Levin DA (2002) The role of chromosomal change in plant evolution. Oxford University Press, Oxford, 230p
- Levy AA, Feldman M (2004) Genetic and epigenetic reprogramming of the wheat genome upon allopolyploidization. *Biol J Linn Soc* 82:607–613
- Lewis KR, John B (1963) Chromosome marker. J & A Churchill, London, p 489
- Lewis KR, John B (1972) The matter of Mendelian heredity, 2nd edn. Longman, London, p 273
- Li C, Zhou A, Sang T (2006) Rice domestication by reducing shattering. *Science* 311:1936–1939
- Li L, Lu K, Chen Z, Mu T, Hu Z, Li X (2008) Dominance, overdominance and epistasis condition the heterosis in two heterotic rice hybrids. *Genetics* 180:1725–1742
- Li T, Liu B, Spalding MH, Weeks DP, Yang B (2012) High-efficiency TALEN-based gene editing produces disease-resistant rice. *Nat Biotechnol* 30:390–392
- Lin CY (1978) Index selection for genetic improvement of quantitative characters. *Theor Appl Genet* 52:49–56
- Lin CS, Binns MR, Lefkovitch LP (1986) Stability analysis: where do we stand? *Crop Sci* 26:894–900
- Lindegren CC (1933) The genetics of *Neurospora*-III. Pure bred stocks and crossing-over in *N crassa*. *Bull Torrey Bot Club* 60:133–154
- Link W, Ghaouti L (2012) Faba bean: breeding for organic farming systems. In: Lammerts van Buren ET, Myers JR (eds) Organic crop breeding. Wiley, Chichester, pp 215–225
- Link W, Ederer W, Metz P, Buiel H, Melchinger AE (2005) Genotypic and environmental variation for degree of cross-fertilization in Faba bean. *Crop Sci* 34:960–964
- Liu BH (1998) Statistical genomics: linkage, mapping, and QTL analysis. CRC, Boca Raton, 611p
- Liu JH, Xu XY, Deng XX (2005) Intergeneric somatic hybridization and its application to crop genetic improvement. *Plant Cell Tissue Organ Cult* 82:19–44
- Liu Z, Liu P, Long F, Hong D, He Q, Yang G (2012) Fine mapping and candidate gene analysis of the nuclear restorer gene Rfp for pol CMS in rapeseed (*Brassica napus* L.). *Theor Appl Genet* 125:773–779. doi:[10.1007/s00122-012-1870-2](https://doi.org/10.1007/s00122-012-1870-2)
- Liu W, Maurer HP, Reif JC, Melchinger AE, Utz HF, Tucker MR, Ranc N, Porta GD, Würschum T (2013) Optimum design of family structure and allocation of resources in association mapping with lines from multiple crosses. *Heredity* 110:71–79
- Love D (1992) Translator's foreword of origin and geography of cultivated plants by N.I. Vavilov. Cambridge University Press, Cambridge
- Lowry DB, Willis JH (2010) A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. *PLOS Biol.* doi:[10.1371/journal.pbio.1000500](https://doi.org/10.1371/journal.pbio.1000500)

- Luna R, Planchon C (1995) Genotype  $\times$  *Bradyrhizobium japonicum* strain interactions in dinitrogen fixation and agronomic traits of soybean (*Glycine max* L. Merr.). *Euphytica* 86:127–134
- Luo ZW, Zhang Z, Leach L, Zhang RM, Bradshaw JE, Kearsey MJ (2006) Constructing genetic linkage maps under a tetrasomic model. *Genetics* 172:2635–2645
- Lusser M, Parisi C, Plan D, Rodríguez-Cerezo E (2012) Deployment of new biotechnologies in plant breeding. *Nat Biotechnol* 30:231–239
- Lv H, Fang Z, Yang L, Zhang Y, Wang Q, Liu Y, Zhuang M, Yang Y, Xie B, Liu B, Liu J, Kang J, Wang X (2014) Mapping and analysis of a novel candidate Fusarium wilt resistance gene *FOC1* in *Brassica oleracea*. *BMC Genomics* 15:1094. doi:[10.1186/1471-2164-15-1094](https://doi.org/10.1186/1471-2164-15-1094)
- Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits. Sinauer, Sunderland, 980p
- Mackay IJ, Gibson JP (1993) The effect of gametic-phase disequilibrium on the prediction of response to recurrent selection in plants. *Theor Appl Genet* 87:152–160
- Maheswaran M, Huang N, Sreerangasamy SR, McCouch SR (2000) Mapping quantitative trait loci associated with days to flowering and photoperiod sensitivity in rice (*Oryza sativa* L.). *Mol Breed* 6:145–155
- Malcolmson JF (1969) Races of *Phytophthora infestans* occurring in Great Britain. *Trans Br Mycol Soc* 53:417–423
- Malthus TR (1798) Malthus: An essay on the principle of population. Cambridge texts in the history of political thought. Paperback 28 Aug 1992. T.R. Malthus (Author), Donald Winch (Editor). Cambridge University Press, Cambridge, 430p
- Mamidi S, Rossi M, Moghaddam SM, Annam D, Lee R, Papa R, McClean PE (2013) Demographic factors shaped diversity in the two gene pools of wild common bean *Phaseolus vulgaris* L. *Heredity* 110:267–276. doi:[10.1038/hdy.2012.82](https://doi.org/10.1038/hdy.2012.82)
- Mannuss A, Trapp O, Puchta H (2012) Gene regulation in response to DNA damage. *Biochim Biophys Acta* 1819:154–165
- Manton I, Sledge WA (1955) Observations on the cytology and taxonomy of the pteridophyte flora of Ceylon. *Phil Trans R Soc B* 238:127–185
- Marcotrigiano M (1997) Chimeras and variegation: patterns of deceit. *HortSci* 32:773–784
- Mardis ER (2008) Next-generation DNA sequencing methods. *Annu Rev Genomics Hum Genet* 9:387–402
- Martin DN, Proebsting WM, Hedden P (1997) Mendel's dwarfing gene: cDNAs from the *Le* alleles and function of the expressed proteins. *Proc Natl Acad Sci* 94:8907–8911
- Martín AC, Shaw P, Phillips D, Reader S, Moore G (2014) Licensing MLH1 sites for crossover during meiosis. *Nat Commun* 5: Article number 4580. doi:[10.1038/ncomms5580](https://doi.org/10.1038/ncomms5580)
- Martínez-Andújar C, Martin RC, Nonogaki H (2012) Seed traits and genes important for translational biology-highlights from recent discoveries. *Plant Cell Physiol* 53:5–15. doi:[10.1093/pcp/pcr112](https://doi.org/10.1093/pcp/pcr112)
- Mather K (1936) Segregation and linkage in autotetraploids. *J Genet* 32:287–314
- Mather KA, Caicedo AL, Polato NR, Olsen KM, McCouch S, Purugganan MD (2007) The extent of linkage disequilibrium in rice (*Oryza sativa* L.). *Genetics* 177:2223–2232
- Maxam AM, Gilbert W (1977) A new method for sequencing DNA. *Proc Natl Acad Sci* 74:560–564
- Maxted N, Kell SP (2009) Establishment of a global network for the *in situ* conservation of crop wild relatives: status and needs. FAO Commission on Genetic Resources for Food and Agriculture, Rome, 266p
- McClintock B (1950) The origin and behaviour of mutable elements in maize. *Proc Natl Acad Sci* 36:344–355
- McGrann GRD, Stavrinides A, Russell J, Corbitt MM, Booth A, Chartrain L, Thomas WTB, Brown JKM (2014) A trade off between *mlo* resistance to powdery mildew and increased susceptibility of barley to a newly important disease, Ramularia leaf spot. *J Exp Bot.* doi:[10.1093/jxb/ert452](https://doi.org/10.1093/jxb/ert452)
- McKevitt S, Ryan T (2013) Project sunshine. Icon Books Ltd, London, 304p

- McKey D, Elias M, Pujol B, Duputie' A (2010) The evolutionary ecology of clonally propagated domesticated plants. *New Phytol* 186:318–332
- McMahon P (2013) Feeding frenzy. Profile Books, London, 314p
- McNaughton IH (1973) Synthesis and sterility of Raphanobrassica. *Euphytica* 22:70–88
- McVey M, Lee SE (2008) MMEJ repair of double-stranded breaks (director's cut): deleted sequences and alternative endings. *Trends Genet* 24:529–538
- McVittie JA, Gale MD, Marshall GA, Westcott B (1978) The intra-chromosomal mapping of the Norin 10 and Tom Thumb dwarfing genes. *Heredity* 40:67–70
- Mead R (1997) Design of plant breeding trials. In: Kempton RA, Fox PN (eds) Statistical methods for plant variety evaluation. Chapman & Hall, London, pp 40–67
- Mead R, Curnow RN, Hasted AM (2002) Statistical methods in agriculture and experimental biology, 3rd edn. Chapman and Hall/CRC, Boca Raton, 488p
- Medina CP (2012) Rice: crop breeding using farmer-led participatory plant breeding. In: Lammerts van Bueren ET, Myers JR (eds) Organic crop breeding. Wiley, Chichester, pp 191–202
- Melchinger AE (2010) The international conference on “heterosis in plants”. *Theor Appl Genet* 120:201–203
- Melgar S, Havey MJ (2010) The dominant *Ms* allele in onion shows reduced penetrance. *J Am Soc Hort Sci* 135:49–52
- Mendel G (1865) Experiments in plant hybridisation. English translation with introduction. (trans: Fisher RA), Oliver & Boyd, Edinburgh, 95p
- Mendiburu AO, Peloquin SJ (1977) The significance of  $2n$  gametes in potato breeding. *Theor Appl Genet* 49:53–61
- Mergoum M, Singh PK, Anderson JA, Peña RJ, Singh RP, Xu SS, Ransom JK (2009a) Spring wheat breeding. In: Carena MJ (ed) Cereals, vol 3, Handbook of plant breeding. Springer, New York, pp 127–156
- Mergoum M, Singh PK, Peña RJ, Río AJL, Cooper KV, Salmon DF, Macpherson HG (2009b) Criticale: a “new” crop with old challenges. In: Carena MJ (ed) Cereals, vol 3, Handbook of plant breeding. Springer, New York, pp 267–287
- Messmer M, Hildermann I, Thorup-Kristensen K, Rengel Z (2012) Nutrient management in organic farming and consequences for direct and indirect selection strategies. In: Lammerts van Bueren ET, Myers JR (eds) Organic crop breeding. Wiley, Chichester, pp 15–38
- Meuwissen TH, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–1829
- Meyer RS, DuVal AE, Jensen HR (2012) Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops. *New Phytol* 196:29–48. doi:[10.1111/j.1469-8137.2012.04253.x](https://doi.org/10.1111/j.1469-8137.2012.04253.x)
- Michaelson-Yeates TPT, Marshall AH, Abberton MT, Rhodes I (1997) Self-compatibility and heterosis in white clover (*Trifolium repens* L.). *Euphytica* 94:341–348
- Miedaner T, Wilde F, Korzun V, Ebmeyer E, Schmolke M, Hartl L, Schön CC (2009) Marker selection for Fusarium head blight resistance based on quantitative trait loci (QTL) from two European sources compared to phenotypic selection in winter wheat. *Euphytica* 166:219–227. doi:[10.1007/s10681-008-9832-0](https://doi.org/10.1007/s10681-008-9832-0)
- Ming R, Moore PH, Wu K-K, D'hont A, Glaszmann JC, Tew TL, Mirkov TE, da Silva J, Jifon J, Rai M, Schnell RJ, Brumley SM, Lakshmanan P, Comstock JC, Paterson AH (2006) Sugarcane improvement through breeding and biotechnology. *Plant Breed Rev* 27:15–118
- Miyao A, Nakagome M, Ohnuma T, Yamagata H, Kanamori H, Katayose Y, Takahashi A, Matsumoto T, Hirochika H (2012) Molecular spectrum of somaclonal variation in regenerated rice revealed by whole-genome sequencing. *Plant Cell Physiol* 53:256–264. doi:[10.1093/pcp/pcr172](https://doi.org/10.1093/pcp/pcr172)
- Moffat A, Wilson JF (2011) The Scots: a genetic journey. Birlinn, Edinburgh, 256p

- Mok DWS, Peloquin SJ (1975) Three mechanisms of  $2n$  pollen formation in diploid potatoes. *Can J Genet Cytol* 17:217–225
- Molina J, Sikora M, Garud N, Flowers JM, Rubinstein S, Reynolds A, Huang P, Jackson S, Schaal BA, Bustamante CD, Boyko AR, Purugganan MD (2011) Molecular evidence for a single evolutionary origin of domesticated rice. *Proc Natl Acad Sci* 108:8351–8356
- Moore R, Clark WD, Vodopich DS (1998) Botany, 2nd edn. WCB/McGraw-Hill, New York, 919p
- Morgan TH, Cattell E (1912) Data for the study of sex-linked inheritance in *Drosophila*. *J Exp Zool* 13:79–101
- Morgante M, De Paoli E, Radovic S (2007) Transposable elements and the plant pan genomes. *Curr Opin Plant Biol* 10:149–155
- Morita R, Kusaba M, Iida S, Yamaguchi H, Nishio T, Nishimura M (2009) Molecular characterization of mutations induced by gamma irradiation in rice. *Genes Genet Syst* 84:361–370
- Morris WL, Ducreux JM, Fraser PD, Millam S, Taylor MA (2006) Engineering ketocarotenoid biosynthesis in potato tubers. *Metab Eng* 8:253–263
- Moseley ME (2001) The Incas and their ancestors: the archaeology of Peru, 2nd edn. Thames and Hudson, London
- Mou B (2008) Lettuce. In: Prohens J, Nuez F (eds) Vegetables I, vol 1, Handbook of plant breeding. Springer, New York, pp 75–116
- Mulema JMK, Olanya OM, Adipala E, Wagoire W (2004/5) Stability of late blight resistance in population B potato clones. *Potato Res* 47: 11–2
- Muller HJ (1916) The mechanism of crossing-over. *Am Nat* 50: 193–221, 284–305, 350–366, 421–434
- Muller HJ (1927) Artificial transmutation of the gene. *Science* 66:84–87
- Muller KO, Black W (1951) Potato breeding for resistance to blight and virus diseases during the last hundred years. *Z Pflanzenzuchung* 31:305–318
- Mundt CC (2014) Durable resistance: a key to sustainable management of pathogens and pests. *Infect Genet Evol* 27:446–455. doi:[10.1016/j.meegid.2014.01.011](https://doi.org/10.1016/j.meegid.2014.01.011), Epub 2014 Jan 31
- Munns R, James RA, Xu B, Athman A, Conn SJ, Jordans C, Byrt CS, Hare RA, Tyerman SD, Tester M, Plett D, Gillham M (2012) Wheat grain yield on saline soils is improved by an ancestral  $\text{Na}^+$  transporter gene. *Nat Biotechnol* 30:360–364. doi:[10.1038/nbt.2120](https://doi.org/10.1038/nbt.2120)
- Murat F, Zhang R, Guizard S, Flores R, Armero A, Pont C, Steinbach D, Quesneville H, Cooke R, Salse J (2014) Shared subgenome dominance following polyploidization explains grass genome evolutionary plasticity from a seven protochromosome ancestor with 16K protogenes. *Genome Biol Evol* 6:12–33. doi:[10.1093/gbe/evt200](https://doi.org/10.1093/gbe/evt200)
- Murphy DJ (2007) People, plants & genes: the story of crops and humanity. Oxford University Press, Oxford, 401p
- Musgrave T, Gardner C, Musgrave W (1998) The plant hunters. Orion, London, 224p
- Myers JR, McKenzie L, Voorrips RE (2012) Brassicas: breeding cole crops for organic agriculture. In: Lammerts van Bueren ET, Myers JR (eds) Organic crop breeding. Wiley, Chichester, pp 251–262
- Myles S, Boyko A, Owens C, Brown P, Grassi F, Aradhya MK, Prins B, Reynolds A, Chi J-M, Ware D, Bustamante CD, Buckler ES (2011) Genetic structure and domestication history of the grape. *Proc Natl Acad Sci* 108:3530–3535
- Naess SK, Bradeen JM, Wielgus SM, Haberlach GT, McGrath JM, Helgeson JP (2000) Resistance to late blight in *Solanum bulbocastanum* is mapped to chromosome 8. *Theor Appl Genet* 101:697–704
- Nagaharu U (1935) Genome analysis in *Brassica* with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. *Jpn J Bot* 7:389–452
- Nally DP (2011) Human encumbrances. University of Notre Dame Press, Notre Dame, 348p
- Newcomb M, Acevedo M, Bockelman HE, Brown-Guedira G, Goates BJ, Jackson EW, Jin Y, Njau P, Rouse MN, Singh D, Wanera R, Bonman JM (2013) Field resistance to the Ug99 race group of the stem rust pathogen in spring wheat landraces. *Plant Dis* 97:882–890

- Newton AC, Hackett C, Swanston JS (2008) Analysing the contribution of component cultivars and cultivar combinations to malting quality, yield and disease in complex mixtures. *J Sci Food Agric* 88:2142–2152
- Niklas KJ (1997) The evolutionary biology of plants. The University of Chicago Press, Chicago, 449p
- Nilan RA, Sideris EG, Kleinhofs A, Sander C, Konzak CF (1973) Azide-a potent mutagen. *Mutat Res* 17:142–144
- Nilsson-Ehle H (1909) Kreuzunguntersuchungen an Hafer und Weizen. *Acta Univ. Lund. Ser* 2, 5, no 2: 1–122
- Oettler G (2005) The fortune of a botanical curiosity—*Triticale*: past, present and future. *J Agric Sci* 143:329–346
- Ogura H (1968) Studies on the new male sterility in Japanese radish, with special references to the utilization of this sterility towards the practical raising of hybrid seeds. *Mem Fac Agric Kagoshima Univ* 6:39–78
- Ollitrault P, Navarro L (2012) Citrus. In: Badenes ML, Byrne DH (eds) *Fruit breeding*, vol 8, *Handbook of plant breeding*. Springer, New York, pp 623–662
- Olsen O, Wang X, von Wettstein D (1993) SA mutagenesis: preferential generation of A-T →G-C transitions in the barley *Ant18* gene. *Proc Natl Acad Sci* 90:8043–8047
- Olsen KM, Sutherland BL, Small LL (2007) Molecular evolution of the *Li/li* chemical defence polymorphism in white clover (*Trifolium repens* L.). *Mol Ecol* 16:4180–4193
- Olsen KM, Hsu SC, Small LL (2008) Evidence on the molecular basis of the *Ac/ac* adaptive cyanogenesis polymorphism in white clover (*Trifolium repens* L.). *Genetics* 179:517–526
- Ooms G, Bossen ME, Burrell MM, Carp A (1986) Genetic manipulation in potato with *Agrobacterium rhizogenes*. *Potato Res* 29:367–379
- Orłowska E, Basile A, Kandzia I, Llorente B, Kirk HG, Cvitanich C (2012) Revealing the importance of meristems and roots for the development of hypersensitive responses and full foliar resistance to *Phytophthora infestans* in the resistant potato cultivar Sarpo Mira. *J Exp Bot* 63:4765–4779
- Ortiz R (1998) Potato breeding via ploidy manipulation. *Plant Breed Rev* 16:15–86
- Ortiz R (2001) The state of the use of potato genetic diversity. In: Cooper HD, Spillane C, Hodgkin T (eds) *Broadening the genetic base of crop production*. CAB International, Wallingford, pp 181–200
- Ortiz R, Huaman Z (1994) Inheritance of morphological and tuber characteristics. In: Bradshaw JE, Mackay GR (eds) *Potato genetics*. CAB International, Wallingford, pp 263–283
- Osman K, Higgins JD, Sanchez-Moran E, Armstrong SJ, Franklin FCH (2011) Pathways to meiotic recombination in *Arabidopsis thaliana*. *New Phytol* 190:523–544. doi:[10.1111/j.1469-8137.2011.03665.x](https://doi.org/10.1111/j.1469-8137.2011.03665.x)
- Owen FV (1945) Cytoplasmically inherited male-sterility in sugar beets. *J Agric Res* 71:423–440
- Owens CL (2008) Grapes. In: Hancock JF (ed) *Temperate fruit crop breeding*. Springer, New York, pp 197–233
- Paine JA, Shipton CA, Chaggar S, Howells RM, Kennedy MJ, Vernon G, Wright SY, Hinchliffe E, Adams JL, Silverstone AL, Drake R (2005) Improving the nutritional value of golden rice through increased pro-vitamin A content. *Nat Biotechnol* 23:482–487. doi:[10.1038/nbt1082](https://doi.org/10.1038/nbt1082)
- Panella L, Lewellen RT (2007) Broadening the genetic base of sugar beet: introgression from wild relatives. *Euphytica* 154:383–400. doi:[10.1007/s10681-006-9209-1](https://doi.org/10.1007/s10681-006-9209-1)
- Panwar DVS, Gupta KR, Battan KR (1991) HKR 228. A semi dwarf aromatic rice strain for Haryana, India. *Int Rice Res News* 16:16–17
- Parisseaux B, Bernardo R (2004) In silico mapping of quantitative trait loci in maize. *Theor Appl Genet* 109:508–514
- Parry B (2004) Trading the genome. Columbia University Press, New York, 319p
- Patel D, Power JB, Anthony P, Badakshi F, Pat Heslop-Harrison JS, Davey MR (2011) Somatic hybrid plants of *Nicotiana × sanderae* (+) *N. debneyi* with fungal resistance to *Peronospora tabacina*. *Ann Bot* 108:809–819. doi:[10.1093/aob/mcr197](https://doi.org/10.1093/aob/mcr197)

- Paterson AH, Bowers JE, Chapman BA (2004) Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. *Proc Natl Acad Sci* 101:9903–9908
- Patterson HD, Hunter EA (1983) The efficiency of incomplete block designs in national list and recommended list cereal variety trials. *J Agric Sci (Camb)* 101:427–433
- Patterson HD, Robinson DL (1989) Row-and-column designs with two replicates. *J Agric Sci (Camb)* 112:73–77
- Patterson HD, Williams ER, Hunter EA (1978) Block designs for variety trials. *J Agric Sci (Camb)* 90:395–400
- Patterson LJ, Wild P, Williams ER (1988) An algorithm to generate designs for variety trials. *J Agric Sci (Camb)* 111:133–136
- Peacock WJ (1970) Replication, recombination and chiasmata in *Goniaea australasiae* (Orthoptera: Acrididae). *Genetics* 65:593–617
- Pearson OH (1972) Cytoplasmically inherited male sterile characters and flavor components from the species *Brassica nigra* (L) Koch × *B. oleracea* L. *J Am Soc Hort Sci Biotechnol* 97:397–402
- Pelletier G, Primard C, Vedel F, Chetrit P (1983) Intergeneric cytoplasmic hybridization in cruciferae by protoplast fusion. *Mol Gen Genet* 191:244–250
- Peloquin SJ, Boiteux LS, Carpoto D (1999) Meiotic mutants in potato: valuable variants. *Genetics* 153:1493–1499
- Pendinen G, Gavrilenko T, Jiang J, Spooner DM (2008) Allopolyploid speciation of the Mexican tetraploid potato species *Solanum stoloniferum* and *S. hirtetingii* revealed by genomic in situ hybridization. *Genome* 51:714–720. doi:[10.1139/G08-052](https://doi.org/10.1139/G08-052)
- Pendinen G, Spooner DM, Jiang J, Gavrilenko T (2012) Genomic in situ hybridization reveals both auto- and allopolyploid origins of different North and Central American hexaploid potato (*Solanum* sect. *Petota*) species. *Genome* 55:407–415
- Peng S, Khush GS (2003) Four decades of breeding for varietal improvement of irrigated lowland rice in the International Rice Research Institute. *Plant Prod Sci* 6:157–164
- Peng J, Richards DE, Hartley NM, Murphy GP, Devos KM, Flintham JE, Beales J, Fish LJ, Worland AJ, Pelica F, Sudhakar D, Christou P, Snape JW, Gale MD, Harberd NP (1999) ‘Green revolution’ genes encode mutant gibberellins response modulators. *Nature* 400:256–261
- Pennisi E (2001) Linnaeus’s last stand? *Science* 291:2304–2307
- Perera ALT, Fahim M, Sriyoheswaran S, Dhanapala MP, Senadhira D, Lawrence MJ (1998) Quantitative genetics of rice I. Evidence of unexploited genetical variation for yield and other quantitative characters in modern Indica cultivars. *Field Crop Res* 55:245–256
- Pérez-Rodríguez P, Riaño-Pachón DM, Corréa LGG, Rensing SA, Kersten B, Mueller-Roeber B (2010) PlnTFDB: updated content and new features of the plant transcription factor database. *Nucleic Acids Res* 38(suppl 1):D822–D827. doi:[10.1093/nar/gkp805](https://doi.org/10.1093/nar/gkp805)
- Perrier X, De Langhe E, Donohue M, Lentfer C, Vrydaghs L, Bakry F, Carreel F, Hippolyte I, Horry J-P, Jenny C, Lebot V, Risterucci A-M, Tomekpe K, Doutreléont H, Ball T, Manwaring J, de Maret P, Denham T (2011) Multidisciplinary perspectives on banana (*Musa* spp.) domestication. *Proc Natl Acad Sci* 108:11311–11318
- Person CO, Sidhu G (1971) In: Proceedings of a panel on mutation breeding for disease resistance, Vienna, IAEA, pp 31–38
- Peteet D (2000) Sensitivity and rapidity of vegetational response to abrupt climate change. *Proc Natl Acad Sci* 97:1359–1361
- Phillips SL, Wolfe MS (2005) Evolutionary plant breeding for low input systems. *J Agric Sci* 143:245–254. doi:[10.1017/S0021859605005009](https://doi.org/10.1017/S0021859605005009)
- Pieterse L, Hils U (2009) World catalogue of potato varieties 2009/10. Agrimedia GmbH, Clenze
- Pillay M, Tripathi L (2007) Banana breeding. In: Kang MS, Priyadarshan PM (eds) Breeding major food staples. Blackwell, Oxford, pp 393–428

- Pingali PL (2012) Green revolution: impacts, limits, and the path ahead. *Proc Natl Acad Sci* 109:12302–12308
- Piperno DR (2001) On maize and the sunflower. *Science* 292:2260–2261
- Piperno DR, Weiss E, Holst I, Nadel D (2004) Processing of wild cereal grains in the Upper Palaeolithic revealed by starch grain analysis. *Nature* 430:670–673
- Piperno DR, Ranere AJ, Holst I, Iriarte J, Dickau R (2009) Starch grain and phytolith evidence for early ninth millennium B.P. maize from the Central Balsas River Valley, Mexico. *Proc Natl Acad Sci* 106:5019–5024
- Plaisted RL (1987) Advances and limitations in the utilization of Neotuberosum in potato breeding. In: Jellis GJ, Richardson DE (eds) *The production of new potato varieties*. Cambridge University Press, Cambridge, pp 186–196
- Plaisted RL, Hoopes RW (1989) The past record and future prospects for the use of exotic potato germplasm. *Am Potato J* 66:603–627
- Pollak LM, Salhuana W (2001) The germplasm enhancement of maize (GEM) project: private and public sector collaboration. In: Cooper HD, Spillane C, Hodgkin T (eds) *Broadening the genetic base of crop production*. CAB International, Wallingford, pp 319–329
- Posselt UK (2010) Breeding methods in cross-pollinated species. In: Boller B, Posselt UK, Veronesi F (eds) *Fodder crops and amenity grasses*, vol 5, *Handbook of plant breeding*. Springer, New York, pp 39–87
- Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 38:904–909
- Primrose SB, Twyman RM (2006) *Principles of gene manipulation and genomics*, 7th edn. Wiley-Blackwell, Oxford, 672p
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
- Prohens J, Nuez F, Carena MJ (2008) *Handbook of plant breeding*. Springer, New York
- Purugganan MD, Fuller DQ (2009) The nature of selection during plant domestication. *Nature* 457:843–848
- Qadir M, Quillérou E, Nangia V, Murtaza G, Singh M, Thomas RJ, Drechsel P, Noble AD (2014) Economics of salt-induced land degradation and restoration. *Nat Res Forum*. doi:[10.1111/1477-8947.12054](https://doi.org/10.1111/1477-8947.12054)
- Qin X, Zhang F, Liu C, Yu H, Cao B, Tian S, Liao Y, Siddique KHM (2015) Wheat yield improvements in China: past trends and future directions. *Field Crop Res* 177:117–124
- Quenouille J, Paulhiac E, Moury B, Palloix A (2014) Quantitative trait loci from the host genetic background modulate the durability of a resistance gene: a rational basis for sustainable resistance breeding in plants. *Heredity* 112:579–587
- Radoev M, Becker HC, Ecke W (2008) Genetic analysis of heterosis for yield and yield components in rapeseed (*Brassica napus* L.) by quantitative trait locus mapping. *Genetics* 179:1547–1558. doi:[10.1534/genetics.108.089680](https://doi.org/10.1534/genetics.108.089680)
- Ramanna MS (1979) A re-examination of the mechanisms of 2*n* gamete formation in potato and its implications for breeding. *Euphytica* 28:537–561
- Ramsay LD, Bradshaw JE, Kearsey MJ (1994) The inheritance of quantitative traits in *Brassica napus* ssp. *rapifera* (swedes): augmented triple test cross analysis of yield. *Heredity* 73:84–91
- Rao GJN, Patnaik A, Chaudhary D (2012) Genetic improvement of Basmati rice through mutation breeding. In: Shu QY, Forster BP, Nakagawa H (eds) *Plant mutation breeding and biotechnology*. CABI, Wallingford, pp 445–452
- Ratcliffe B, El-Dien OG, Kláspět J, Porth I, Chen C, Jaquish B, El-Kassaby YA (2015) A comparison of genomic selection models across time in Interior Spruce (*Picea engelmannii* × *glauca*) using high density SNP imputation methods. *Heredity* advance online publication, July 1, 2015; doi:[10.1038/hdy.2015.57](https://doi.org/10.1038/hdy.2015.57)
- Raun WR, Johnson GV (1999) Improving nitrogen use efficiency for cereal production. *Agron J* 91:357–363

- Ravel C, Charmet G (1996) A comprehensive multisite recurrent selection strategy in perennial ryegrass. *Euphytica* 88:215–226
- Rawat DS, Anand IJ (1979) Male sterility in Indian mustard. *Indian J Genet Plant Breed* 39:412–414
- Ray DK, Mueller ND, West PC, Foley JA (2013) Yield trends are insufficient to double global crop production by 2050. *PLoS One* 8(6), e66428. doi:[10.1371/journal.pone.0066428](https://doi.org/10.1371/journal.pone.0066428)
- Reader J (2008) Propitious esculent. William Heinemann, London, 315p
- Reddy PS, Thammiraju B (1977) Inheritance of fertility restoration in sunflower (*Helianthus annuus* L.). *Euphytica* 26:409–412
- Reddy NRR, Ragimasaalwada M, Sabbavarapu MM, Nadoor S, Patil JV (2014) Detection and validation of stay-green QTL in post-rainy sorghum involving widely adapted cultivar, M35-1 and a popular stay-green genotype B35. *BMC Genomics* 15:909. doi:[10.1186/1471-2164-15-909](https://doi.org/10.1186/1471-2164-15-909)
- Reheul D, De Cauwer B, Cougnon M (2010) The role of forage crops in multifunctional agriculture. In: Boller B, Posselt UK, Veronesi F (eds) *Fodder crops and amenity grasses*, vol 5, *Handbook of plant breeding*. Springer, New York, pp 1–12
- Reid TA, Yang RC, Salmon DF, Navabi A, Spaner D (2011) Realized gains from selection for spring wheat grain yield are different in conventional and organically managed systems. *Euphytica* 177:253–266
- Reinstädler A, Müller J, Czembor JH, Piffanelli P, Panstruga R (2010) Novel induced *mlo* mutant alleles in combination with site-directed mutagenesis reveal functionally important domains in the heptahelical barley *Mlo* protein. *BMC Plant Biol* 10:31. doi:[10.1186/1471-2229-10-31](https://doi.org/10.1186/1471-2229-10-31)
- Reisch BI, Owens CL, Cousins PS (2012) Grape. In: Badenes ML, Byrne DH (eds) *Fruit breeding*, vol 8, *Handbook of plant breeding*. Springer, New York, pp 225–262
- Rhoades MM (1950) Meiosis in maize. *J Hered* 41:58–67
- Riddit WM (1925) The turnip crop. In: Paterson WGR (ed) *Farm crops*, vol II, *Root crops*. Gresham, London, pp 69–118
- Riechmann JL, Heard J, Martin G, Reuber L, Jiang C, Keddie J, Adam L, Pineda O, Ratcliffe OJ, Samaha RR, Creelman R, Pilgrim M, Broun P, Zhang JZ, Ghandehari D, Sherman BK, Yu G (2000) Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. *Science* 290:2105–2110
- Rietman H, Bijsterbosch G, Cano LM, Lee HR, Vossen JH, Jacobsen E, Visser RGF, Kamoun S, Vleeshouwers VGAA (2012) Qualitative and quantitative late blight resistance in the potato cultivar Sarpo Mira is determined by the perception of five distinct RXLR effectors. *Mol Plant Microbe Interact* 25:910–919
- Riley R, Chapman V (1958) Genetic control of the cytological diploid behaviour of hexaploid wheat. *Nature* 182:713–715
- Riley R, Chapman V, Johnson R (1968) Introduction of yellow rust resistance of *Aegilops comosa* into wheat by genetically induced homoeologous recombination. *Nature* 217:383–384. doi:[10.1038/217383a0](https://doi.org/10.1038/217383a0)
- Rios D, Ghislain M, Rodriguez F, Spooner DM (2007) What is the origin of the European potato? Evidence from Canary Island landraces. *Crop Sci* 47:1271–1280
- Ritland K (2002) Extensions of models for the estimation of mating systems using *n* independent loci. *Heredity* 88:221–228. doi:[10.1038/sj.hdy.6800029](https://doi.org/10.1038/sj.hdy.6800029)
- Roach BT (1995) Sugar canes. In: Smartt J, Simmonds NW (eds) *Evolution of crop plants*, 2nd edn. Longman Scientific & Technical, Harlow, pp 160–166
- Robb W, Wishart R (1915) The improvement of the feeding value of swedes and turnips. University of St Andrews, Agricultural Department. University Press, St Andrews, pp 1–12
- Roberts EH (1991) Genetic conservation in seed banks. *Biol J Linn Soc* 43:23–29
- Robertson A (1960) A theory of limits in artificial selection. *Proc R Soc B* 153:234–249
- Robin M-M (2008) The world according to Monsanto. The New Press, New York, 372p
- Robinson DL (1984) A study of sequential variety selection systems. *J Agric Sci (Camb)* 102:119–126

- Robinson DL, Kershaw CD, Ellis RP (1988) An investigation of two-dimensional yield variability in breeders' small plot trials. *J Agri Sci (Camb)* 111:419–426
- Rodríguez GR, Muños S, Anderson C, Sim S-C, Michel A, Causse M, McSpadden Gardener BB, Francis D, van der Knaap E (2011) Distribution of SUN, OVATE, LC, and FAS in the tomato germplasm and the relationship to fruit shape diversity. *Plant Physiol* 156:275–285
- Rogers JS, Edwardson JR (1952) The utilization of cytoplasmic male sterility inbreds in the production of corn hybrids. *Agron J* 44:8–12
- Rommens CM, Ye J, Richael C, Swords K (2006) Improving potato storage and processing characteristics through all-native DNA transformation. *J Agric Food Chem* 54:9882–9887
- Rommens CM, Haring MA, Swords K, Davies HV, Belknap WR (2007) The intragenic approach as a new extension to traditional plant breeding. *Trends Plant Sci* 12:397–403
- Rommens CM, Yan H, Swords K, Richael C, Ye J (2008) Low-acrylamide French fries and potato chips. *Plant Biotechnol J* 6:843–853
- Rooney W (2012) Sorghum breeding. In: Acquaah G (ed) *Principles of plant genetics and breeding*. Wiley-Blackwell, Chichester, pp 620–625
- Rosenthal DM, Ort DR (2012) Examining cassava's potential to enhance food security under climate change. *Trop Plant Biol* 5:30–38
- Rosenthal DM, Slattery RA, Miller RE, Grennan AK, Cavagnaro TR, Fauquet CM, Gleadow RM, Ort DR (2012) Cassava about-FACE: Greater than expected yield stimulation of cassava (*Manihot esculenta*) by future CO<sub>2</sub> levels. *Glob Chang Biol* 18:2661–2675. doi:[10.1111/j.1365-2486.2012.02726.x](https://doi.org/10.1111/j.1365-2486.2012.02726.x)
- Ross H (1986) Potato breeding—problems and perspectives, vol 13, *Advances in plant breeding*. Paul Parey, Berlin
- Rotherham ID (2013) *The lost fens: England's greatest ecological disaster*. The History Press, Stroud, 207p
- Rouppé van der Voort JNAM, van der Vossen E, Bakker E, Overmars H, van Zandvoort P, Hutten R, Lankhorst RK, Bakker J (2000) Two additive QTL conferring broad-spectrum resistance in potato to *Globodera pallida* are localized on resistance gene clusters. *Theor Appl Genet* 101:1122–1130
- Rousseau-Gueutin M, Lerceteau-Köhler E, Barrot L, Sargent DJ, Monfort A, Simpson D, Arús P, Guérin G, Denoyes-Rothan B (2008) Comparative genetic mapping between octoploid and diploid *Fragaria* species reveals a high level of colinearity between their genomes and the essentially disomic behavior of the cultivated octoploid strawberry. *Genetics* 179:2045–2060. doi:[10.1534/genetics.107.083840](https://doi.org/10.1534/genetics.107.083840)
- Roy SJ, Negrão S, Tester M (2014) Salt resistant crop plants. *Curr Opin Biotechnol* 26:115–124
- Royal Botanic Gardens, Kew (1998) *Plants+People: an exhibition of items from the Economic Botany Collections*. Pardoe Blacker, Lingfield, 64p
- Russell JR, Ellis RP, Thomas WTB, Waugh R, Provan J, Booth A, Fuller J, Lawrence P, Young G, Powell W (2000) A retrospective analysis of spring barley germplasm development from “foundation genotypes” to currently successful cultivars. *Mol Breed* 6:533–568
- Ryuto H, Fukunishi N, Hayashi Y, Ichida H, Abe T, Kase M, Yano Y (2008) Heavy-ion beam irradiation facility for biological samples in RIKEN. *Plant Biotechnol* 25:119–122
- Salaman RN (1910) The inheritance of colour and other characters in the potato. *J Genet* 1:7–46
- Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain terminating inhibitors. *Proc Natl Acad Sci* 74:5463–5467
- Sansaloni CP, Petroli CD, Carling J, Hudson CJ, Steane DA, Myburg AA, Grattapaglia D, Vaillancourt RE, Kilian A (2010) A high-density diversity arrays technology (DArT) microarray for genome-wide genotyping in *Eucalyptus*. *Plant Methods* 6:16. doi:[10.1186/1746-4811-6-16](https://doi.org/10.1186/1746-4811-6-16)
- Sardos J, McKey D, Duval MF, Malapa R, Noyer JL, Lebot V (2008) Evolution of cassava (*Manihot esculenta* Crantz) after recent introduction into a South Pacific Island system: the contribution of sex to the diversification of a clonally propagated crop. *Genome* 51:912–921

- Sato Y, Shirasawa K, Takahashi Y, Nishimura M, Nishio T (2006) Mutant selection of progeny of gamma-ray-irradiated rice by DNA heteroduplex cleavage using *Brassica* petiole extract. Breed Sci 56:179–183
- Sato Y, Morita R, Nishimura M, Yamaguchi H, Kusaba M (2007) Mendel's green cotyledon gene encodes a positive regulator of the chlorophyll-degrading pathway. Proc Natl Acad Sci 104:14169–14174. doi:[10.1073/pnas.0705521104](https://doi.org/10.1073/pnas.0705521104)
- Savitsky VF (1950) Monogerm sugar beets in the United States. Proc ASSBT 6:156–159
- Scarelli N, Tostain S, Vigouroux Y, Agbangla C, Daïnou O, Pham J-L (2006) Farmers' use of wild relative and sexual reproduction in a vegetatively propagated crop. The case of yam in Benin. Mol Ecol 15:2421–2431
- Schmidt A (1924) Histologische studien an phanerogamen vegetationspunkten. Bot Arch 7 (8):345–404
- Schnable JC, Springer NM, Freeling M (2011) Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. Proc Natl Acad Sci 108:4069–4074
- Schneider A, Molnár I, Molnár-Láng M (2008) Utilisation of *Aegilops* (goatgrass) species to widen the genetic diversity of cultivated wheat. Euphytica 163:1–19. doi:[10.1007/s10681-007-9624-y](https://doi.org/10.1007/s10681-007-9624-y)
- Scholten OE, van Heusden AW, Khrustaleva LI, Burger-Meijer K, Mank RA, Antonise RGC, Harrewijn JL, Van Haecke W, Oost EH, Peters RJ, Kik C (2007) The long and winding road leading to the successful introgression of downy mildew resistance into onion. Euphytica 156:345–353. doi:[10.1007/s10681-007-9383-9](https://doi.org/10.1007/s10681-007-9383-9)
- Schweizer D (1973) Differential staining of plant chromosomes with Giemsa. Chromosoma 40:307–320
- Scoretti KC, Creste S, Calsa T, Xavier MA, Landell MGA, Figueira A, Benedito VA (2012) Challenges, opportunities and recent advances in sugarcane breeding. In: Abdurakhmonov IY (ed) Plant breeding. InTech Europe, Rijeka, pp 267–296. doi:[10.5772/1389](https://doi.org/10.5772/1389)
- Scriber JM (2001) *Bt* or not *Bt*: is that the question? Proc Natl Acad Sci 98:12328–12330
- Scurrall M, Celis-Gamboa C, Chumbiauca S, Salas A, Visser RGF (2008) Hybridization between wild and cultivated potato species in the Peruvian Andes and biosafety implications for deployment of GM potatoes. Euphytica 164:881–892
- Searle SR (1966) Matrix algebra for the biological sciences. Wiley, New York, 296p
- Sears ER (1956) The transfer of leaf rust resistance from *Aegilops umbellulata* to wheat. "Genetics in plant breeding". In: Brookhaven Symposia in Biology 9, pp 1–22
- Semel Y, Nissenbaum J, Menda N, Zinder M, Krieger U, Issman N, Pleban T, Lippman Z, Gur A, Zamir D (2006) Overdominant quantitative trait loci for yield and fitness in tomato. Proc Natl Acad Sci 103:12981–12986
- Seppelt R, Manceur AM, Liu J, Fenichel EP, Klotz S (2014) Synchronized peak-rate years of global resources use. Ecol Soc 19:50. doi:[10.5751/ES-07039-190450](https://doi.org/10.5751/ES-07039-190450)
- Sham P (1998) Statistics in human genetics. Arnold, London, 290p
- Shaner G, Finney RE (1977) The effect of nitrogen fertilization on the expression of slow-mildewing resistance in Knox wheat. Phytopathology 67:1051–1056
- Shigyo M, Kik C (2008) Onion. In: Prohens J, Nuez F (eds) Vegetables II, vol 2, Handbook of plant breeding. Springer, New York, pp 121–159
- Shikazono N, Suzuki C, Kitamura S, Watanabe H, Tano S, Tanaka A (2005) Analysis of mutations induced by carbon ions in *Arabidopsis thaliana*. J Exp Bot 56:587–596
- Shu QY, Forster BP, Nakagawa H (2012) Plant mutation breeding and biotechnology. CABI, Wallingford, 608p
- Shull GH (1908) The composition of a field of maize. Am Breeders' Assoc Rep 4:296–301
- Shull GH (1909) A pure-line method of corn breeding. Am Breed Assoc Rep 5:51–59
- Sigareva MA, Earle ED (1997) Direct transfer of a cold-tolerant Ogura male-sterile cytoplasm into cabbage (*Brassica oleracea* ssp. *capitata*) via protoplast fusion. Theor Appl Genet 94:213–220

- Sim S-C, Durstewitz G, Plieske J, Wieseke R, Ganal MW, Deynze AV, Hamilton JP, Buell CR, Causse M, Wijeratne S, Francis DM (2012) Development of a large SNP genotyping array and generation of high-density genetic maps in tomato. *PLoS One* 7(7), e40563. doi:[10.1371/journal.pone.0040563](https://doi.org/10.1371/journal.pone.0040563)
- Simillion C, Vandepoele K, Van Montagu MCE, Zabeau M, Van de Peer Y (2002) The hidden duplication past of *Arabidopsis thaliana*. *Proc Natl Acad Sci* 99:13627–13632
- Simmonds NW (1969) Prospects of potato improvement. Scottish Plant Breeding Station Forty-Eighth Annual Report 1968–69, pp 18–38
- Simmonds NW (1981) Genotype (*G*), Environment (*E*) and *GE* components of crop yields. *Exp Agric* 17:355–362
- Simmonds NW (1991) Genetics of horizontal resistance to diseases of crops. *Biol Rev* 66:189–241
- Simmonds NW (1995a) Potatoes. In: Smartt J, Simmonds NW (eds) Evolution of crop plants, 2nd edn. Longman Scientific & Technical, Harlow, pp 466–471
- Simmonds NW (1995b) Bananas. In: Smartt J, Simmonds NW (eds) Evolution of crop plants, 2nd edn. Longman Scientific & Technical, Harlow, pp 370–375
- Simmonds NW (1996) Family selection in plant breeding. *Euphytica* 90:201–208
- Simmonds NW, Wastie RL (1987) Assessment of horizontal resistance to late blight of potatoes. *Ann Appl Biol* 111:213–221
- Simon PW, Freeman RE, Vieira JV, Boiteux LS, Briard M, Nothnagel T, Michalik B, Kwon Y-S (2008) Carrot. In: Prohens J, Nuez F (eds) Vegetables II, vol 2, Handbook of plant breeding. Springer, New York, pp 327–357
- Simons KJ, Fellers JP, Trick HN, Zhang Z, Tai Y-S, Gill BS, Faris JD (2006) Molecular characterization of the major wheat domestication gene *Q*. *Genetics* 172:547–555
- Singh RP, Trethowan R (2007) Breeding spring bread wheat for irrigated and rainfed production systems of the developing world. (2007). In: Kang MS, Priyadarshan PM (eds) Breeding major food staples. Blackwell, Oxford, pp 109–140
- Singh RP, Hodson DP, Huerta-Espino J, Jin Y, Bhavani S, Njau P, Herrera-Foessel S, Singh PK, Singh S, Govindan V (2011) The emergence of Ug99 races of the stem rust fungus is a threat to world wheat production. *Annu Rev Phytopathol* 49:465–481
- Sinnott EW, Dunn LC, Dobzhansky T (1958) Principles of genetics, 5th edn. McGraw-Hill, New York, 459p
- Skibbe DS, Schnable PS (2005) Male sterility in maize. *Maydica* 50:367–376
- Skoglund P, Malmström H, Omrak A, Raghavan M, Valdiosera C, Günther T, Hall P, Tambets K, Parik J, Sjögren K-G, Apel J, Willerslev E, Storå J, Götherström A, Jakobsson M (2014) Genomic diversity and admixture differs for Stone-Age Scandinavian foragers and farmers. *Science*. doi:[10.1126/science.1253448](https://doi.org/10.1126/science.1253448)
- Skøt L, Humphreys J, Humphreys MO, Thorogood D, Gallagher J, Sanderson R, Armstead IP, Thomas ID (2007) Association of candidate genes with flowering time and water-soluble carbohydrate content in *Lolium perenne* (L.). *Genetics* 177:535–547
- Slater A, Scott NW, Fowler MR (2008) Plant biotechnology: the genetic manipulation of plants, 2nd edn. Oxford University Press, Oxford, 376p
- Sleper DA, Poehlman JM (2006) Breeding field crops, 5th edn. Blackwell, Oxford, 424p
- Smartt J, Simmonds NW (1995) Evolution of crop plants, 2nd edn. Longman Scientific & Technical, Harlow, 531p
- Smith HF (1936) A discriminant function for plant selection. *Ann Eugen* 7:240–250
- Smith BD (1989) Origins of agriculture in eastern North America. *Science* 246:1566–1571
- Smith DR, Hooker AL, Lim SM, Beckett JB (1971) Disease reaction of thirty sources of cytoplasmic male-sterile corn to *Helminthosporium maydis* race T. *Crop Sci* 11:772–773
- Sofi PA, Rather AG, Wani SA (2007) Genetic and molecular basis of cytoplasmic male sterility in maize. *Commun Biometry Crop Sci* 2:49–60
- Soh AC, Wong CK, Ho YW, Choong CW (2009) Oil Palm. In: Vollmann J, Rajcan I (eds) Oil crops, vol 4, Handbook of plant breeding. Springer, New York, pp 333–367

- Sokal RR, Oden NL, Wilson C (1991) Genetic evidence for the spread of agriculture in Europe by demic diffusion. *Nature* 351:143–145
- Solomon-Blackburn RM, Stewart HE, Bradshaw JE (2007) Distinguishing major-gene from field resistance to late blight (*Phytophthora infestans*) of potato (*Solanum tuberosum*) and selecting for high levels of field resistance. *Theor Appl Genet* 115:141–149
- Sood S, Kuraparthi V, Bai G, Gill BS (2009) The major threshability genes soft glume (*sog*) and tenacious glume (*Tg*), of diploid and polyploid wheat, trace their origin to independent mutations at non-orthologous loci. *Theor Appl Genet* 119:341–351
- Spielmeyer W, Ellis MH, Chandler PM (2002) Semidwarf (*sd-1*), “green revolution” rice, contains a defective gibberellin 20 oxidase gene. *Proc Natl Acad Sci* 99:9043–9048
- Spillane C, Gepts P (2001) Evolutionary and genetic perspectives on the dynamics of crop genepools. In: Cooper HD, Spillane C, Hodgkin T (eds) Broadening the genetic base of crop production. CAB International, Wallingford, pp 25–70
- Spooner DM, Huijmans RJ (2001) Potato systematics and germplasm collecting, 1989–2000. *Am J Potato Res* 78:237–268
- Spooner DM, Salas A (2006) Structure, biosystematics, and genetic resources. In: Gopal J, Khurana SMP (eds) Handbook of potato production, improvement, and postharvest management. Food Products Press, New York, pp 1–39
- Spooner DM, Hettterscheid WLA, van den Berg RG, Brandenburg WA (2003) Plant nomenclature and taxonomy. *Hortic Rev* 28:1–60
- Spooner DM, van den Berg RG, Rodriguez A, Bamberg J, Huijmans RJ, Lara-Cabrera SI (2004) Wild potatoes (*Solanum* section *Petota*; Solanaceae) of North and Central America. *Syst Bot Monogr* 68:1–209. doi:[10.2307/25027915](https://doi.org/10.2307/25027915)
- Spooner DM, McLean K, Ramsay G, Waugh R, Bryan GJ (2005) A single domestication for potato based on multilocus amplified fragment length polymorphism genotyping. *Proc Natl Acad Sci U S A* 102:14694–14699
- Springer NM, Ying K, Fu Y, Ji T, Yeh C-T, Jia Y, Wu W, Richmond T, Kitzman J, Rosenbaum H, Iniguez AL, Barbazuk WB, Jeddelloh JA (2009) Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content. *PLoS Genet* 5, e1000734
- Stadler LJ (1926) The variability of crossing-over in maize. *Genetics* 11:1–37
- Stadler LJ (1928a) Genetic effects of X rays in maize. *Proc Natl Acad Sci U S A* 14:69–75
- Stadler LJ (1928b) Mutations in barley induced by X-rays and radium. *Science* 68:186–187
- Stam P (1993) Construction of integrated genetic linkage maps by means of a new computer package: JOINMAP. *Plant J* 3:739–744
- Stephens JC, Holland PF (1954) Cytoplasmic male sterility for hybrid sorghum seed production. *Agron J* 46:20–23
- Stewart HE, Wastie RL (1989) A rapid scoring technique for potato tuber disease assessment. *Potato Res* 32:353–357
- Stewart HE, Bradshaw JE, Wastie RL (1994) Correlation between resistance to late blight in foliage and tubers in potato clones from parents of contrasting resistance. *Potato Res* 37:429–434
- Stewart HE, Bradshaw JE, Pande B (2003) The effect of the presence of R-genes for resistance to late blight (*Phytophthora infestans*) of potato (*Solanum tuberosum*) on the underlying level of field resistance. *Plant Pathol* 52:193–198
- Stich B, Mohring J, Piepho H-P, Heckenberger M, Buckler ES, Melchinger AE (2008) Comparison of mixed-model approaches for association mapping. *Genetics* 178:1745–1754
- Stiekema WJ, Heidekamp F, Louwerse JD, Verhoeven HA, Dijkhuis P (1988) Introduction of foreign genes into potato cultivars Bintje and Desiree using an *Agrobacterium tumefaciens* binary vector. *Plant Cell Rep* 7:47–50
- Storck T, Böhme T, Schultheiss H (2012) Fortuna et al. Status and perspectives of GM approaches to fight late blight. In: Thirteenth EuroBlight workshop St. Petersburg (Russia), 9–12 Oct 2011. PPO-Special Report no. 15, pp 45–48

- Struik PC, Lommen WJM, Haverkort AJ, Storey RMJ (2007) The canon of potato science. *Potato Res* 50
- Sturtevant AH (1913) The linear arrangement of six sex-linked factors in *Drosophila*, as shown by their mode of association. *J Exp Zool* 14:43–59
- Sukhotu T, Hosaka K (2006) Origin and evolution of Andigena potatoes revealed by chloroplast and nuclear DNA markers. *Genome* 49:636–647
- Sun G, Zhu C, Kramer MH, Yang S-S, Song W, Piepho H-P, Yu J (2010) Variation explained in mixed-model association mapping. *Heredity* 105:333–340
- Suneson CA (1951) Male-sterile facilitated synthetic hybrid barley. *Agron J* 43:234–236
- Suneson CA (1956) An evolutionary plant breeding method. *Agron J* 48:188–191
- Suprasanna P, Jain SM, Ochatt SJ, Vm K, Predieri S (2012) In: Shu QY, Forster BP, Nakagawa H (eds) *Plant mutation breeding and biotechnology*. CABI, Wallingford, pp 371–385
- Sutton WS (1903) The chromosomes in heredity. *Biol Bull* 4:231–248
- Swaminathan MS, Howard HW (1953) The cytology and genetics of the potato (*Solanum tuberosum*) and related species. *Bibliogr Genet* 16:1–192
- Swanson CP, Merz T, Young WJ (1967) *Cytogenetics*. Prentice-Hall, New Jersey, 194p
- Świeżynski KM, Domański L, Zarzycka H, Zimnoch-Guzowska E (2000) The reaction of potato differentials to *Phytophthora infestans* isolates collected in nature. *Plant Breed* 119:119–126
- Szarejko I (2012) Haploid mutagenesis. In: Shu QY, Forster BP, Nakagawa H (eds) *Plant mutation breeding and biotechnology*. CABI, Wallingford, pp 387–410
- Tai GCC (1994) Use of  $2n$  gametes. In: Bradshaw JE, Mackay GR (eds) *Potato genetics*. CAB International, Wallingford, pp 109–132
- Takagi Y, Anai T, Nakagawa H (2012) Mutation breeding for fatty acid composition in soybean. In: Shu QY, Forster BP, Nakagawa H (eds) *Plant mutation breeding and biotechnology*. CABI, Wallingford, pp 437–444
- Takagi H, Abe A, Yoshida K, Kosugi S, Natsume S, Mitsuoka C, Uemura A, Utsushi H, Tamiru M, Takuno S, Innan H, Cano LM, Kamoun S, Terauchi R (2013) QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. *Plant J* 74:174–183. doi:[10.1111/tpj.12105](https://doi.org/10.1111/tpj.12105)
- Talbot M (1984) Yield variability of crop varieties in the U.K. *J Agric Sci (Camb)* 102:315–321
- Talbot M (1997) Resource allocation for selection systems. In: Kempton RA, Fox PN (eds) *Statistical methods for plant variety evaluation*. Chapman & Hall, London, pp 162–174
- Talbot M, Milner AD, Nutkins MAE, Law JR (1995) The effect of interference between plots on yield performance in crop variety trials. *J Agric Sci (Camb)* 124:335–342
- Tallury SP, Goodman MM (2001) The state of the use of maize genetic diversity in the USA and Sub-Saharan Africa. In: Cooper HD, Spillane C, Hodgkin T (eds) *Broadening the genetic base of crop production*. CAB International, Wallingford, pp 159–179
- Tan S, Bowe SJ (2012) Herbicide-tolerant crops developed from mutations. In: Shu QY, Forster BP, Nakagawa H (eds) *Plant mutation breeding and biotechnology*. CABI, Wallingford, pp 423–436
- Tan S, Evans RR, Dahmer ML, Singh BK, Shaner DL (2005) Imidazolinone-tolerant crops: history, current status and future. *Pest Manag Sci* 61:246–257. doi:[10.1002/ps.993](https://doi.org/10.1002/ps.993)
- Tanaka A, Shikazono N, Hase Y (2010) Studies on biological effects of ion beams on lethality, molecular nature of mutation, mutation rate, and spectrum of mutation phenotype for mutation breeding in higher plants. *J Radiat Res* 51:223–233
- Tanaka Y, Tsuda M, Yasumoto K, Yamagishi H, Terachi T (2012) A complete mitochondrial genome sequence of Ogura-type male-sterile cytoplasm and its comparative analysis with that of normal cytoplasm in radish (*Raphanus sativus* L.). *BMC Genomics* 13:352
- Tang H, Cuevas HE, Das S, Sezen UU, Zhou C, Guo H, Goff VH, Ge Z, Clemente TE, Paterson AH (2013) Seed shattering in a wild sorghum is conferred by a locus unrelated to domestication. *Proc Natl Acad Sci* 110:15824–15829

- Tarn TR, Tai GCC (1983) *Tuberosum × Tuberosum* and *Tuberosum × Andigena* potato hybrids: comparisons of families and parents, and breeding strategies for Andigena potatoes in long-day temperate environments. *Theor Appl Genet* 66:87–91
- Tatum LA (1971) The southern corn leaf blight epidemic. *Science* 171:1113–1116
- Taylor JH (1953) Autoradiographic detection of incorporation of P<sup>32</sup> into chromosomes during meiosis and mitosis. *Exp Cell Res* 4:164–173
- Taylor JH (1965) Distribution of tritium-labeled DNA among chromosomes during meiosis. I. Spermatogenesis in the grasshopper. *J Cell Biol* 25:57–67
- Taylor JH, Woods PS, Hughes WL (1957) The organization and duplication of chromosomes as revealed by autoradiographic studies using tritium-labeled thymidine. *Proc Natl Acad Sci* 43:122–128
- Teixeira JEC, Weldekidan T, de Leon N, Flint-Garcia S, Holland JB, Lauter N, Murray SC, Xu W, Hessel DA, Kleintop AE, Hawk JA, Hallauer A, Wisser RJ (2015) Hallauer's Tusón: a decade of selection for tropical-to-temperate phenological adaptation in maize. *Heredity* 114:229–240
- Tek AL, Stevenson WR, Helgeson JP, Jiang J (2004) Transfer of tuber soft rot and early blight resistances from *Solanum brevidens* into cultivated potato. *Theor Appl Genet* 109:249–254
- Thomas WTB (1987) The use of random F<sub>3</sub> families for cross prediction in spring barley. *J Agric Sci* 108:431–436
- Thompson KF (1957) Self-incompatibility in marrow-stem kale, *Brassica oleracea* var *acephala*. I. Demonstration of a sporophytic system. *J Genet* 55:45–60
- Thompson KF (1964) Triple-cross hybrid kale. *Euphytica* 13:173–177
- Thompson KF (1967) Breeding problems in kale (*Brassica oleracea*) with particular reference to marrow-stem kale. *Rep Pl Breed Inst*, Cambridge, 1965–66, pp 7–34
- Thompson KF, Taylor J (1967) The breakdown of self-incompatibility in cultivars of *Brassica oleracea*. *Heredity* 21:637–648
- Tibshirani R (1996) Regression shrinkage and selection via the LASSO. *J R Stat Soc Ser B* 58:267–288
- Till BJ, Reynolds SH, Weil C, Springer N, Burtner C, Young K, Bowers E, Codomo CA, Enns LC, Odden AR, Greene EA, Comai L, Henikoff S (2004) Discovery of induced point mutations in maize genes by TILLING. *BMC Plant Biol* 4:12. doi:[10.1186/1471-2229-4-12](https://doi.org/10.1186/1471-2229-4-12)
- Till BJ, Cooper J, Tai TH, Colowit P, Greene EA, Henikoff S, Comai L (2007) Discovery of chemically induced mutations in rice by TILLING. *BMC Plant Biol* 7:19. doi:[10.1186/1471-2229-7-19](https://doi.org/10.1186/1471-2229-7-19)
- Tomczynska I, Stefanczyk E, Chmielarz M, Karasiewicz B, Kaminski P, Jones JD, Lees AK, Sliwka J (2014) A locus conferring effective late blight resistance in potato cultivar Sarpo Mira maps to chromosome XI. *Theor Appl Genet* 127:647–657
- Toms EM, Cornish MA, Kearsey MJ (1994a) Comparison of family and sib selection during selfing. *Heredity* 73:635–641
- Toms EM, Kearsey MJ, Cornish MA (1994b) Correlated response to selection during selfing. *Heredity* 73:642–649
- Trognitz BR, Bonierbale M, Landeo JA, Forbes G, Bradshaw JE, Mackay GR, Waugh R, Huarte MA, Colon L (2001) Improving potato resistance to disease under the global initiative on late blight. In: Cooper HD, Spillane C, Hodgkin T (eds) *Broadening the genetic base of crop production*. CAB International, Wallingford, pp 385–398
- Troyer AF (2004) Champaign County, Illinois, and the origin of hybrid corn. In: Janick J (ed) *Plant breeding reviews*, vol 24, Part 1: Long-term selection: maize. Wiley, Hoboken, pp 41–59
- Troyer AF (2006) Adaptedness and heterosis in corn and mule hybrids. *Crop Sci* 46:528–543
- Uauy C, Paraiso F, Colasuonno P, Tran RK, Tsai H, Berardi S, Comai L, Dubcovsky J (2009) A modified TILLING approach to detect induced mutations in tetraploid and hexaploid wheat. *BMC Plant Biol* 9:115. doi:[10.1186/1471-2229-9-115](https://doi.org/10.1186/1471-2229-9-115)
- Ullstrup AJ (1972) The impacts of the southern corn leaf blight epidemic of 1970–71. *Annu Rev Phytopathology* 10:37–50

- Utz HF, Melchinger AE, Schön CC (2000) Bias and sampling error of the estimated proportion of genotypic variance explained by quantitative trait loci determined from experimental data in maize using cross validation and validation with independent samples. *Genetics* 154:1839–1849
- Utz HF, Bohn M, Melchinger AE (2001) Predicting progeny means and variances of winter wheat crosses from phenotypic values of their parents. *Crop Sci* 41:1470–1478
- Van der Plank JE (1968) Disease resistance in plants. Academic, New York, 206p
- Van Ginkel M, Ortiz-Monasterio I, Trethowan R, Hernandez E (2001) Methodology for selecting segregating populations for improved N-use efficiency in bread wheat. *Euphytica* 119:223–230
- van Harten AM (1998) Mutation breeding theory and applications. Cambridge University Press, Cambridge, 353p
- van Heerwaarden J, Doebley J, Briggs WH, Glaubitz JC, Goodman MM, Gonzalez JJS, Ross-Ibarra J (2011) Genetic signals of origin, spread, and introgression in a large sample of maize landraces. *Proc Natl Acad Sci* 108:1088–1092
- van Heerwaarden J, Hufford MB, Ross-Ibarra J (2012) Historical genomics of North American maize. *Proc Natl Acad Sci* 109:12420–12425
- Van Oeveren AJ (1993) Efficiency of single seed descent and early selection in the breeding of self-fertilizing crops. PhD thesis, University of Wageningen
- Van Oeveren AJ, Stam P (1992) Comparative simulation studies on the effects of selection for quantitative traits in autogamous crops: early selection versus single seed descent. *Heredity* 69:342–351
- Van Ooijen JW (2006) JoinMap 4, Software for the calculation of genetic linkage maps in experimental populations. Kyazma B.V, Wageningen
- Van Ooijen JW, Jansen J (2013) Genetic mapping in experimental populations. Cambridge University Press, Cambridge, 157p
- Vanholme R, Demedts B, Morreel K, Ralph J, Boerjan W (2010) Lignin biosynthesis and structure. *Plant Physiol* 153:895–905. doi:[10.1104/pp.110.155119](https://doi.org/10.1104/pp.110.155119)
- Vanuytrecht E, Raes D, Willems P, Geerts S (2012) Quantifying field-scale effects of elevated carbon dioxide concentration on crops. *Climate Res* 54:35–47. doi:[10.3354/cr01096](https://doi.org/10.3354/cr01096)
- Vaughan DA, Lu B-R, Tomooka N (2008) The evolving story of rice evolution. *Plant Sci* 174:394–408
- Vavilov NI (1940) The theory of the origin of cultivated plants after Darwin. Nauka [Science] 2:55–75
- Veilleux RE (1985) Diploid and polyploid gametes in crop plants: mechanisms of formation and utilization in plant breeding. *Plant Breed Rev* 3:253–288
- Veilleux RE (2005) Cell and tissue culture of potato (Solanaceae). In: Razdan MK, Mattoo AK (eds) Genetic improvement of Solanaceous crops, vol I, Potato. Science, Enfield, pp 185–208
- Velasco R, Zharkikh A, Affourtit J, Dhingra A, Cestaro A, Kalyanaraman A, Fontana P, Bhatnagar SK, Troggio M et al (2010) The genome of the domesticated apple (*Malus × domestica* Borkh.). *Nat Genet* 42:833–839
- Verde I, Abbott AG, Scalabrin S, Jung S, Shu S, Marroni F et al (2013) The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. *Nat Genet* 45:487–494
- Veronesi F, Brummer EC, Huyghe C (2010) Alfalfa. In: Boller B, Posselet UK, Veronesi F (eds) Fodder crops and amenity grasses, vol 5, Handbook of plant breeding. Springer, New York, pp 395–437
- Virmani SS, Ilyas-Ahmed M (2007) Rice breeding for sustainable production. In: Kang MS, Priyadarshan PM (eds) Breeding major food staples. Blackwell, Oxford, pp 141–191
- Visa S, Cao C, Gardener BM, van der Knaap E (2014) Modeling of tomato fruits into nine shape categories using elliptic Fourier shape modeling and Bayesian classification of contour morphometric data. *Euphytica* 200:429–439. doi:[10.1007/s10681-014-1179-0](https://doi.org/10.1007/s10681-014-1179-0)
- Vollman J, Menken M (2012) Soybean: breeding for organic farming systems. In: Lammerts van Bueren ET, Myers JR (eds) Organic crop breeding. Wiley, Chichester, pp 203–214

- von Caemmerer S, Quick WP, Furbank RT (2012) The development of C<sub>4</sub> rice: current progress and future challenges. *Science* 336:1671–1672. doi:[10.1126/science.1220177](https://doi.org/10.1126/science.1220177)
- Walker AR, Lee E, Bogs J, McDavid DA, Thomas MR, Robinson SP (2007) White grapes arose through the mutation of two similar and adjacent regulatory genes. *Plant J* 49:772–785
- Walley PG, Teakle GR, Moore JD, Allender CJ, Pink DAC, Buchanan-Wollaston V, Barker GC (2012) Developing genetic resources for pre-breeding in *Brassica oleracea* L.: an overview of the UK perspective. *J Plant Biotechnol* 39:62–68. doi:[10.5010/JPB.2012.39.1.062](https://doi.org/10.5010/JPB.2012.39.1.062)
- Walsh B (2004) Population- and quantitative-genetic models of selection limits. In: Janick J (ed) *Plant breeding reviews*, vol 24, Part 1: Long-term selection maize. Wiley, Hoboken, pp 177–225
- Walters TW, Mutschler MA, Earle ED (1992) Protoplast fusion-derived Ogura male sterile cauliflower with cold tolerance. *Plant Cell Rep* 10:624–628
- Wang H, Nussbaum-Wagler T, Li B, Zhao Q, Vigouroux Y, Faller M, Bomblies K, Lukens L, Doebley JF (2005) The origin of the naked grains of maize. *Nature* 436:714–719. doi:[10.1038/nature03863](https://doi.org/10.1038/nature03863)
- Wang X, Shen J, Liao H (2010) Acquisition or utilization, which is more critical for enhancing phosphorus efficiency in modern crops? *Plant Sci* 179:302–306
- Wang H-Y, Liu ZH, Chen P-D, Wang X-E (2012) Irradiation-facilitated chromosomal translocation: wheat as an example. In: Shu QY, Forster BP, Nakagawa H (eds) *Plant mutation breeding and biotechnology*. CABI, Wallingford, pp 223–239
- Wang N, Long T, Yao W, Xiong L, Zhang Q, Wu C (2013) Mutant resources for the functional analysis of the rice genome. *Mol Plant* 6:596–604. doi:[10.1093/mp/sss142](https://doi.org/10.1093/mp/sss142)
- Wang C-H, Zheng X-M, Xu Q, Yuan X-P, Huang L, Zhou H-F, Wei X-H, Ge S (2014a) Genetic diversity and classification of *Oryza sativa* with emphasis on Chinese rice germplasm. *Heredity* 112:489–496. doi:[10.1038/hdy.2013.130](https://doi.org/10.1038/hdy.2013.130)
- Wang Y, Cheng X, Shan Q, Zhang Y, Liu J, Gao C, Qiu J-L (2014b) Simultaneous editing of three homeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nat Biotechnol* 32:947–951. doi:[10.1038/nbt.2969](https://doi.org/10.1038/nbt.2969)
- Warthmann N, Chen H, Ossowski S, Weigel D, Hervé P (2008) Highly specific gene silencing by artificial miRNAs in rice. *PLoS One* 3(3), e1829. doi:[10.1371/journal.pone.0001829](https://doi.org/10.1371/journal.pone.0001829)
- Wastie RL, Caligari PDS, Wale SJ (1988a) Assessing the resistance of potatoes to powdery scab [*Spongospora subterranea* (Wallr.) Lagerh.]. *Potato Res* 31:167–171
- Wastie RL, Caligari PDS, Stewart HE, Mackay GR (1988b) Assessing the resistance to gangrene of progenies of potato (*Solanum tuberosum* L.) from parents differing in susceptibility. *Potato Res* 31:355–365
- Waterworth WM, Drury GE, Bray CM, West CE (2011) Repairing breaks in the plant genome: the importance of keeping it together. *New Phytol* 192:805–822
- Watson JD (2003) DNA: the secret of life. William Heinemann, London, 446p
- Watson JD, Crick FHC (1953a) A structure for deoxyribose nucleic acid. *Nature* 171:737–738
- Watson JD, Crick FHC (1953b) Genetical implications of the structure of deoxyribose nucleic acid. *Nature* 171:964–967
- Watts LE (1970) Productivity of F<sub>1</sub> hybrids of botanical varieties of *Brassica oleracea* L. *Euphytica* 19:398–404
- Weir BS (1996) Genetic data analysis II. Sinauer, Sunderland, 445p
- Weir BS, Cockerham CC (1973) Mixed selfing and random mating at two loci. *Genet Res* 21:247–261
- Weiss E, Wetterstrom W, Nadel D, Bar-Yosef O (2004) The broad spectrum revisited: evidence from plant remains. *Proc Natl Acad Sci* 101:9551–9555
- Welch JE, Grimball EL (1947) Male sterility in carrot. *Science* 106:594
- Wenzel G (1994) Tissue culture. In: Bradshaw JE, Mackay GR (eds) *Potato genetics*. CAB International, Wallingford, pp 173–195
- Werner JE, Peloquin SJ (1990) Inheritance and two mechanisms of 2n egg formation in 2x potatoes. *J Hered* 81:371–374

- Whipple CJ, Kebrom TH, Weber AL, Yang F, Hall D, Meeley R, Schmidt R, Doebley J, Brutnell TP, Jackson DP (2011) *grassy tillers1* promotes apical dominance in maize and responds to shade signals in the grasses. Proc Natl Acad Sci 108(33):E506–E512
- White S, Shaw D (2009) The usefulness of late-blight resistant Sárpo cultivars—a case study. Acta Hortic 834:161–166
- Whitehorn PR, O'Connor S, Wackers FL, Goulson D (2012) Neonicotinoid pesticide reduces bumble bee colony growth and queen production. Science 336:351–352. doi:[10.1126/science.1215025](https://doi.org/10.1126/science.1215025)
- Whitehouse HLK (1973) Towards an understanding of the mechanism of heredity, 3rd edn. Edward Arnold, London, 445p
- Wijnker E, Deurhof L, van de Belt J, de Snoo CB, Blankestijn H, Becker F, Ravi M, Chan SWL, van Dun K, Lelivelt CLC, de Jong H, Dirks R, Joost JB Keurentjes (2014) Hybrid recreation by reverse breeding in *Arabidopsis thaliana*. Nat Protocols 9:761–772. doi:[10.1038/nprot.2014.049](https://doi.org/10.1038/nprot.2014.049)
- Willis KJ, McElwain JC (2014) The evolution of plants, 2nd edn. Oxford University Press, Oxford, 398p
- Wilson A (1993) The story of the potato through illustrated varieties. Balding and Mansell, England, 120p
- Witcombe JR, Virk DS (2001) Number of crosses and population size for participatory and classical plant breeding. Euphytica 122:451–462
- Wolyn DJ, Chahal A (1998) Nuclear and cytoplasmic interactions for petaloid male-sterile accessions of wild carrot (*Daucus carota* L.). J Am Soc Hort Sci 123:849–853
- Wood C, Habgood N (2010) Why people need plants. Royal Botanic Gardens, Kew, 192 p
- Wricke G, Weber WE (1986) Quantitative genetics and selection in plant breeding. Walter de Gruyter, Berlin, 406p
- Wright AJ (1979) The use of differential coefficients in the development and interpretation of quantitative genetic models. Heredity 43:1–8
- Wright AJ, Cockerham CC (1985) Selection with partial selfing. I. Mass selection. Genetics 109:585–597
- Wright AJ, Cockerham CC (1986) Selection with partial selfing. II. Family selection. Crop Sci 26:261–268
- Wu F, Tanksley SD (2010) Chromosomal evolution in the plant family Solanaceae. BMC Genomics 11:182. <http://www.biomedcentral.com/1471-2164/11/182>
- Würschum T, Kraft T (2014) Cross-validation in association mapping and its relevance for the estimation of QTL parameters of complex traits. Heredity 112:463–468
- Würschum T, Kraft T (2015) Evaluation of multi-locus models for genome-wide association studies: a case study in sugar beet. Heredity 114:281–290. doi:[10.1038/hdy.2014.98](https://doi.org/10.1038/hdy.2014.98)
- Xu S (2003) Theoretical basis of the Beavis effect. Genetics 165:2259–2268
- Xu Y (2010) Molecular plant breeding. CABI, Wallingford, 734p
- Xu G, Fan X, Miller AJ (2012) Plant nitrogen assimilation and use efficiency. Annu Rev Plant Biol 63:153–182. doi:[10.1146/annurev-arplant-042811-105532](https://doi.org/10.1146/annurev-arplant-042811-105532)
- Yan L, Loukoianov A, Tranquilli G, Helguera M, Fahima T, Dubcovsky J (2003) Positional cloning of the wheat vernalization gene *VRN1*. Proc Natl Acad Sci 100:6263–6268
- Yan L, Loukoianov A, Blechl A, Tranquilli G, Ramakrishna W, SanMiguel P, Bennetzen JL, Echenique V, Dubcovsky J (2004) The wheat *VRN2* gene is a flowering repressor down-regulated by vernalization. Science 303:1640–1644
- Yan L, Fu D, Li C, Blechl A, Tranquilli G, Bonafede M, Sanchez A, Valarik M, Yasuda S, Dubcovsky J (2006) The wheat and barley vernalization gene *VRN3* is an orthologue of *FT*. Proc Natl Acad Sci 103:19581–19586
- Yan W, Kang MS, Ma B, Woods S, Cornelius PL (2007) GGE biplot vs. AMMI analysis of genotype-by-environment data. Crop Sci 47:643–653
- Yang Z (2014) Molecular evolution: a statistical approach. Oxford University Press, Oxford, 512p

- Yang D, Zhang Y, Zhu Z, Chen T, Zhao Q, Yao S, Zhao L, Zhu W, Wang C (2013a) Substitutional mapping the cooked rice elongation by using chromosome segment substitution lines in rice. *Mol Plant Breed* 4:107–115. doi:[10.5376/mpb.2013.04.0013](https://doi.org/10.5376/mpb.2013.04.0013)
- Yang Q, Li Z, Li W, Ku L, Wang C, Ye J, Ye J, Li K, Yang N, Li Y, Zhong T, Li J, Chen Y, Yan J, Yang X, Xu M (2013b) CACTA-like transposable element in ZmCCT attenuated photoperiod sensitivity and accelerated the postdomestication spread of maize. *Proc Natl Acad Sci* 110:16969–16974
- Yates F (1936) A new method of arranging variety trials involving a large number of varieties. *J Agric Science (Camb)* 26:424–455
- Yates F (1937) A further note on the arrangement of variety trials: Quasi-Latin squares. *Ann Eugen* 7:121–140
- Ye X, Al-Babili S, Klöti A, Zhang J, Lucca P, Beyer P, Potrykus I (2000) Engineering the provitamin A ( $\beta$ -Carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. *Science* 287:303–305. doi:[10.1126/science.287.5451.303](https://doi.org/10.1126/science.287.5451.303)
- Yoshida K, Schuenemann VJ, Cano LM, Pais M, Mishra B, Sharma R, Lanz C, Martin FN, Kamoun S, Krause J, Thines M, Weigel D, Burbano HA (2013) The rise and fall of the *Phytophthora infestans* lineage that triggered the Irish potato famine. *eLife* 2, e00731
- Young SSY (1961) A further examination of the relative efficiency of three methods of selection under less-restricted conditions. *Genet Res* 2:106–121
- Yu J, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES (2005) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet* 38:203–208
- Zeng ZB (1993) Theoretical basis of separation of multiple linked gene effects on mapping quantitative trait loci. *Proc Natl Acad Sci* 90:10972–10976
- Zeng ZB (1994) Precision mapping of quantitative trait loci. *Genetics* 136:1457–1466
- Zeng ZB, Wang T, Zou W (2005) Modeling quantitative trait loci and interpretation of models. *Genetics* 169:1711–1725
- Zhang NW, Pelgrom K, Niks RE, Visser RGF, Jeuken MJW (2009a) Three combined quantitative trait loci from nonhost *Lactuca saligna* are sufficient to provide complete resistance of lettuce against *Bremia lactucae*. *Mol Plant Microbe Interact* 22:1160–1168
- Zhang X, Pérez-Rodríguez P, Kassa S, Beyene Y, Babu R, López Cruz MA, Vicente FS, Olsen M, Buckler E, Jannink J-L, Prasanna BM, Crossa J (2015) Genomic prediction in bi-parental tropical maize populations in water-stressed and well-watered environments using low density and GBS SNPs. *Heredity* 114:291–299
- Zhou G, Chen Y, Yao W, Zhang C, Xie W, Hua J, Xing Y, Xiao J, Zhang Q (2012) Genetic composition of yield heterosis in an elite rice hybrid. *Proc Natl Acad Sci* 109:15847–15852
- Zhu Y, Chen H, Fan J, Wang Y, Li Y, Chen J, Fan JX, Yang S, Hu L, Leungk H, Mewk TW, Tengk PS, Wangk Z, Mundtk CC (2000) Genetic diversity and disease control in rice. *Nature* 406:718–722
- Zhu BF, Si L, Wang Z, Zhou Y, Zhu J, Shangguan Y, Lu D, Fan D, Li C, Lin H, Qian Q, Sang T, Zhou B, Minobe Y, Han B (2011) Genetic control of a transition from black to straw-white seed hull in rice domestication. *Plant Physiol* 155:1301–1311. doi:[10.1104/pp.110.168500](https://doi.org/10.1104/pp.110.168500)
- Zhu QH, Upadhyaya N, Helliwell C (2012) Transposon mutagenesis for functional genomics. In: Shu QY, Forster BP, Nakagawa H (eds) *Plant mutation breeding and biotechnology*. CABI, Wallingford, pp 507–522
- Ziółkowski PA, Kaczmarek M, Babula D, Sadowski J (2006) Genome evolution in *Arabidopsis/Brassica*: conservation and divergence of ancient rearranged segments and their breakpoints. *Plant J* 47:63–74
- Zobel RW, Wright MJ, Gauch HG Jr (1988) Statistical analysis of a yield trial. *Agron J* 80:388–393

- Zuellig MP, Kenney AM, Sweigart AL (2014) Evolutionary genetics of plant adaptation: insights from new model systems. *Curr Opin Plant Biol* 18:44–50
- Zurn JD, Newcomb M, Rouse MN, Jin Y, Chao S, Sthapit J, See DR, Wanyera R, Njau P, Bonman JM, Brueggeman R, Acevedo M (2014) High-density mapping of a resistance gene to Ug99 from the Iranian landrace PI 626573. *Mol Breed* 34:871–881. doi:[10.1007/s11032-014-0081-8](https://doi.org/10.1007/s11032-014-0081-8)

# Index

## A

Abiotic stresses, 59, 71, 190, 208, 260, 263, 493, 497, 559, 561, 580, 591  
Abu Hureyra. *See* Farming  
Adaptation, 17, 21, 25, 61, 65, 67, 141, 146, 148, 150, 158, 190, 207, 208, 212, 219, 228–230, 263, 279, 445, 483, 484, 520, 522, 624  
Adaptations, to new environments, 17, 21, 148–151  
*Aegilops comosa*, 517  
*Aegilops speltoides*. *See* Wheat B genome  
*Aegilops tauschii*. *See* Wheat D genome  
*Aegilops umbellulata*. *See* *Triticum umbellulata*  
*Agrobacterium*-mediated transformation. *See* Genetic transformation  
*Agrobacterium*-mediated transformation, potato, 564  
*Agrobacterium rhizogenes*, 564  
*Agrobacterium tumefaciens*, 45, 553, 554, 563, 564, 575  
AGROBASE Generation II™, 168  
*Agropyron elongatum*, 519  
Aleurone layer, 82, 117  
Alfalfa (*Medicago sativa*), 309  
Alfalfa, anthracnose (*Colletotrichum meliloti*), 310  
Alfalfa, HarvXtra™, 586  
Alfalfa, low lignin, 586  
Alfalfa, stem nematode (*Ditylenchus dipsaci*), 310  
*Allium cepa*. *See* Onions  
*Allium roylei*, 506, 507

Aluminium tolerance, 221  
*Amblyopyrum muticum*, 498  
Amici, Giovanni, 30  
Angiosperm Phylogeny Group III. *See* Plant classification  
Angiosperms. *See* Flowering plants  
Anthers, 30, 33, 35, 80, 81, 84, 110, 378, 402, 416, 420, 443, 476  
Anther culture, 501  
Anthocyanins, 23, 152  
Apomixis, 344, 385  
Apomixis, clonal cultivars, 385, 386  
Apple (*Malus × domestica*), 383  
Apple, powdery mildew (*Podosphaera leucotricha*), 383  
Apple, scab (*Venturia inaequalis*), 383  
Apple breeding, 382, 383  
Apple breeding, genomic selection, 384  
Aquifer, the Ogallala, 7  
*Arabidopsis thaliana*, 43, 140, 258, 259, 544, 546, 559, 623  
*Arabidopsis thaliana*, CONSTANS gene for flowering time, 192  
*Arabidopsis thaliana*, *FLOWERING LOCUS T (FT)*, 151  
*Arabidopsis thaliana*, Gibberellin Insensitive gene (GAI), 156  
*Arabidopsis–Brassica*, 258–259  
*Arachis duranensis*. *See* Peanut, wild  
*Arachis hypogaea*. *See* Peanut, cultivated  
*Arachis ipaënsis*. *See* Peanut, wild  
Ascomycete fungi. *See* Fungi  
Australian Plant Phenomics Facility, 176

**B**

- Bacillus thuringiensis*, 577, 582  
*Bacillus thuringiensis*, Cry endotoxin genes, 577  
 Bacterial resistance, 622  
 Bacterial resistance, broad spectrum, 616, 622–623  
 Bailey, Liberty Hyde, 33  
 Banana wilt (*Fusarium oxysporum* f.sp. *cubense* (Panama disease)), 256  
 Bananas (*Musa spp.*), 255, 372  
 Bananas, wild (*Musa acuminata*), 497  
 Bananas, wild, *Musa balbisiana*, 497  
 Barley (*Hordeum vulgare*), 275, 497, 621  
 Barley, Mlo powdery mildew resistance, 621  
 Barley, powdery mildew (*Blumeria graminis* f.sp. *hordei*), 621  
 Barley, Ramularia leaf spot (*Ramularia collo-cygni*), 622  
 Barley, six-rowed, 83, 146, 147  
 Barley, two-rowed, 83, 146, 147  
 Base broadening, 213  
 Base broadening, incorporation, 45, 520, 526  
 Base broadening, maize, 520  
 Base broadening, *Neotuberosum*, 521  
 Base broadening, potatos, 520–522  
 BASF Plant Science, 582, 583  
 Basmati rice breeding programme, 531, 542  
 Bateson, William, 115, 133  
 Beal, William, 33, 292  
 Bean, common (*Phaseolus vulgaris*), 13, 17, 40, 586  
 Best Linear Unbiased Estimators (BLUEs). *See* General mixed model  
 Best Linear Unbiased Predictors (BLUPs). *See* General mixed model  
*Beta vulgaris*. *See* Sugar and fodder beet  
 Biffen, Rowland, 596  
 Binomial distribution, 111–113, 120, 121, 164, 187, 188, 346, 348, 351, 352, 407, 431  
 Biodiversity International, 47, 50, 485  
 Biofortification, 54  
 Biofuels, 58, 59  
   second generation, 373  
 Biogas production, 59  
 Biolistic-mediated transformation. *See* Genetic transformation  
*Bipolaris maydis*. *See* Maize, southern corn leaf blight  
*Blumeria graminis* f.sp. *hordei*. *See* Barley, powdery mildew  
 Borlaug Global Rust Initiative, 619

**Borlaug, Norman**, 47, 48, 619

- Boyer, Herb**, 43  
*Bradyrhizobium japonicum* (rhizobial strains), 208  
*Brassica carinata*. *See* Mustard  
*Brassica juncea*. *See* Mustard  
*Brassica napus*. *See* Oilseed rape and swedes  
*Brassica nigra*. *See* Mustard  
*Brassica oleracea*. *See* Broccoli, Brussels sprouts, Cabbage and Kale  
*Brassica oleracea*, Anand CMS from *B. tournefortii*, 417  
*Brassica oleracea*, botanical varieties, 413  
*Brassica oleracea*, cytoplasmic male sterility (CMS), 416  
*Brassica oleracea*, Ogura CMS from radish, 416  
*Brassica oleracea*, self-incompatibility, 414  
*Brassica rapa*, 36, 241, 258, 259, 417, 477  
   *See also* Turnip  
*Brassica tournefortii*, 417  
 Brassicas, clubroot (*Plasmodiophora brassicae*), 293  
*Bremia lactucae*, 507  
 Broad bean (*Vicia faba*), 86, 118, 277, 340  
 Broccoli (*Brassica oleracea*), 294, 413, 417  
 Brown, Robert, 37  
 Brussels sprouts (*Brassica oleracea*), 258, 413, 415  
 Bryophytes, 78  
 Bulk population selection, 426, 444–445, 455  
 Bulmer effect, 313, 314

**C**

- Cabbage (*Brassica oleracea*), 258, 413, 415, 417, 604  
 Cabbage, Fusarium wilt (*Fusarium oxysporum* f. sp. *conglutinans*), 604  
*Camellia sinensis*. *See* Tea  
 Camerarius, Rudolph, 30  
 Canola. *See* Oilseed rape  
*Capsicum annuum*. *See* Pepper  
*Carica papaya*. *See* Papaya  
 Carrot (*Daucus carota*), 24, 420  
 Carrot, cytoplasmic male sterility (CMS), 420–422  
 Carrot, F<sub>1</sub> hybrids, 420  
 Carrot, petaloid CMS, 420  
 Carrot breeding, 420, 421  
 Carson, Rachel, 70  
*Cassava* (*Manihot esculenta*), 19, 278, 497  
 Cassava breeding, 370  
 Cassava Mosaic Disease (CMD), 369, 370

- Cassava*, bacterial blight (*Xanthomonas axonopodis*), 369  
Cell divisions, 37, 38, 43, 75, 76, 78, 80, 81, 84, 85, 93, 111, 240, 243, 533  
Cell divisions, cytokinesis, 86, 243  
Cell lineages, 81–82  
Cell lineages, chimeras, 81–82  
Cell nucleus, 86  
Cellular basis of life, 36–37  
Central Potato Research Institute of India, 522  
Centres of origin of cultivated plants, 8, 484  
*Cercospora beticola*. *See* Sugar beet, leaf spot  
CGIAR, 492, 520  
*Chenopodium quinoa*. *See* Quinoa  
Chimeric cultivars, 81  
Chimeric structure, 82  
Chimeric structure, dissociation, 533  
*Chi-square* ( $\chi^2$ ) goodness-of-fit statistic, 113–115  
Chloroplast genome, 233, 240–241, 566  
Chloroplasts, 31, 37, 138, 242, 243, 404, 416, 477, 500, 575  
Chromosomes, aneuploids, 92, 385  
Chromosomes, behaviour, 38, 111, 242  
Chromosomes, centric fusion, 92  
Chromosomes, centromere, 38, 86, 87, 92, 101, 111, 118, 125, 244, 247, 250, 253, 518  
Chromosomes, changes in number, 92  
Chromosomes, chiasma, 87, 247  
Chromosomes, disjunction, 87, 91  
Chromosomes, evolution in the Solanaceae, 105–106  
Chromosomes, Giemsa staining, 84  
Chromosomes, heterochromatin, 92, 104  
Chromosomes, homologous, 443, 510, 511, 515  
Chromosomes, karyotype, 84  
Chromosomes, number, 38, 92, 127, 257, 260, 264, 265, 286, 385, 422, 443, 499, 503, 508  
Chromosomes, pairing (synapsis), 87, 89  
Chromosomes, preparations, 84  
Chromosomes, rearrangements in wheat, 104–106  
Chromosomes, reciprocal translocations, 104, 105, 518  
Chromosomes, sister chromatids, 85, 87, 88, 249, 250  
Chromosomes, staining, 84  
Chromosomes, supernumerary B, 92  
Chromosomes, synaptonemal complex, 89  
*Citrus clementina* x *C. tangerina*. *See* Mandarins  
*Citrus* species, 243, 257–258  
Clark, James, 33  
Climate change, 62–64  
Climate change, rising carbon dioxide concentrations, 63, 360  
Climate change, rising temperatures, 67  
Climate, global, 4  
Climate, Younger Dryas Interval, 4  
*Clivia cyrtanthiflora*, 85  
Clover, red (*Trifolium pratense*), 241, 310  
Clover, white (*Trifolium repens*), 135, 311  
*Coffea arabica*. *See* Coffee, arabica  
*Coffea canephora*. *See* Coffee, robusta  
Coffee, arabica (*Coffea arabica*), 20, 496  
Coffee, robusta (*Coffea canephora*), 20  
Cohen, Stanley, 43  
Coke, Thomas, 26  
Colchicine, use of, 84, 241, 257, 310  
Cold shock proteins, 580  
Cold tolerance, 416, 547  
*Colletotrichum meliloti*. *See* Alfalfa, anthracnose  
Combining ability, 410, 412  
Combining ability, general, 331, 335, 338, 340, 363, 370, 396–398, 401  
Combining ability, specific, 335, 363, 397–399, 407, 410  
Commonwealth Potato Collection (CPC), 361, 485, 490, 491, 521  
Composite crosses, 444, 520  
Composite interval mapping.  
    *See* QTL analysis  
Computer simulation, 186, 253–255, 294, 312, 316, 329, 411, 450, 451, 453–455, 514  
Conservation  
    Conservation, collecting germplasm, 492–493  
    Conservation, core collection size, 493  
    Conservation, core collections, 493–495  
    Conservation, cryopreservation, 495  
    Conservation, effect of climate change, 489  
    Conservation, *ex situ*, 490–491  
    Conservation, gene management zones, 490  
    Conservation, germplasm sample size, 492–493  
    Conservation, *in situ*, 488–490  
    Conservation, landraces, 488, 489  
    Conservation, protected areas, 488  
    Conservation, storage of seeds, 490

- Consultative Group for International Agricultural Research (CGIAR), 50, 486
- Contingency table, 114, 115
- Continuous variation, 40–41, 161, 164
- Convention on Biological Diversity (CBD), 485, 487
- Cooked rice length (CRL), 190
- Cotton (*Gossypium spp.*), 577
- Cotton, bollworm (*Helicoverpa armigera*), 578
- Cotton, herbicide tolerant, 52
- Cotton, insect resistant, 52, 53, 577, 578
- Crassulacean acid metabolism (CAM), 64
- Crick, Francis, 43, 44
- CRISPR/Cas9. *See* DNA sequence modification
- Crossability, 32, 498, 500–503, 508
- Crossbreeding, 71, 286, 387–390, 392–397, 399–402, 404, 406–410, 412, 414–424
- Crosses
- Crosses, discarding at  $F_1$  and  $F_3$ , 433–439
  - Crosses, initial choice, 431–433
  - Crosses, number to make, 426
  - Crosses, progeny sizes, 426, 439
  - Cross-validation, 185, 202–205
  - Cultivar registration, 36
  - Cultivars
  - Cultivars, clonal, 281, 343–346, 350, 352–355, 357–379, 381–386, 445, 455
  - Cultivars, hybrids. *See* Hybrid cultivars
  - Cultivars, inbred line, 49, 50, 289, 312, 395, 431, 455, 460, 472–474, 478
  - Cultivars, mixtures. *See* Mixtures
  - Cultivars, open-pollinated, 47, 280, 387, 390, 391, 393–395, 402, 409, 414, 417, 421, 422, 459, 475, 478, 479, 523
  - Cultivars, synthetic. *See* Synthetic cultivars
  - Cusco Potato Park, 489
  - Cyanogenesis, 134, 135, 311
  - Cybrid, 499
  - Cytological techniques, 84–85
  - Cytoplasmic male sterility (CMS), 47, 402, 403, 411, 416, 417, 421, 473, 475, 478, 479, 496, 497
- D**
- Dactylis glomerata* spp. *glomerata*, 296
  - Daktulosphaira vitifoliae*. *See* Grapevines, phylloxera root louse
  - Darwin, Charles, 22, 31, 34–35
  - Datura innoxia*, 443
  - Datura stramonium*, 92
  - Daucus carota*, 24, 420
- de Moivre, Abraham, 164
- de Saussure, Nicholas Theodore, 63
- Diabrotica virgifera virgifera*. *See* Maize, Western corn rootworm
- Diallel. *See* mating design
- Diazotrophic endophytes, 373
- Disease nurseries, 613–615
- Disease progress curve, 608–615
- Disease resistance
- Disease resistance, breeding, 592–594
  - Disease resistance, broad spectrum, 622–623
  - Disease resistance, gene-for-gene concept, 596–603
  - Disease resistance, general, 595
  - Disease resistance, horizontal resistance, 603–605
  - Disease resistance, host differentials, 600
  - Disease resistance, oligogenic, 596–603
  - Disease resistance, qualitative, 595, 603
  - Disease resistance, quantitative, 595, 603
  - Disease resistance, receptor-effector concept, 599
  - Disease resistance, receptors, 599, 600
  - Disease resistance, *R*-genes from *Solanum demissum*, 594, 600–602
  - Disease resistance, screening, 605–608
  - Disease resistance, specific, 602, 615, 622
  - Disease resistance, types, 595–596
  - Disease resistance, vertical, 603–605
  - Dispersion of cultivated plants, 15–20
  - Dispersion of cultivated plants, New and Old Worlds, 16–19
  - Ditylenchus dipsaci*. *See* Alfalfa, Stem nematode
  - Diversity Fixed Foundation Sets, 495
  - DNA damage, 93–95, 97, 533
  - DNA, double helix, 43, 75, 76, 93
  - DNA, replication, 86, 88, 94, 96, 97, 100
  - DNA, transcription, 93, 94, 136
  - DNA repair
    - DNA repair, double strand break, 89, 91, 95–96, 541, 543
    - DNA repair, excision, 95
    - DNA repair, micro-homology-mediated end joining (MMEJ), 96
    - DNA repair, non homologous end joining (NHEJ), 95
    - DNA repair, photoreactivation, 94–95
    - DNA repair, synthesis-dependent strand annealing (SDSA), 96  - DNA replication, 545
  - DNA sequence modification

- DNA sequence modification, CRISPR/Cas9 system, 557
- DNA sequence modification, engineered nucleases, 555
- DNA sequence modification, gene editing, 556
- DNA sequence modification, gene insertion, 556
- DNA sequence modification, gene replacement, 556
- DNA sequence modification, mutagenesis, 555
- DNA sequence modification, site-directed, 554
- DNA sequence modification, structural changes, 557
- DNA sequencing, 539, 544, 599
- DNA sequencing, next-generation, 77
- Domestication
- Domestication, centres of, 8–9
  - Domestication, debates, 11–14
  - Domestication, features, 10–11
  - Domestication, number of genes, 147
  - Domestication, syndrome, 10
  - Domestication, traits, 143, 146, 147
- Domestication genes
- Domestication genes, black hull in rice (*Bh4*), 22, 144
  - Domestication genes, free threshing in wheat (*Q*), 145
  - Domestication genes, glume tenacity in wheat (*Tg*), 146
  - Domestication genes, grassy tillers<sub>1</sub> in maize (*gt1*), 143
  - Domestication genes, shattering in sorghum (*Sh1*), 144
  - Domestication genes, shattering QTL in rice (*qSH1*), 144
  - Domestication genes, *shattering4* in rice (*sh4*), 22, 143
  - Domestication genes, size of lateral spikelets in barley (*Int-c*), 147
  - Domestication genes, soft glume in wheat (*sog*), 146
  - Domestication genes, *teosinte branched1* in maize (*tb1*), 143
  - Domestication genes, *teosinte glume architecture* in maize (*tga1*), 143
  - Domestication genes, two-rowed spike in barley (*Vrs1*), 146
- Double reduction, 249
- Doubled haploids (DH), 92, 93, 200, 231, 287, 417, 418, 426, 443, 446–448, 455, 476, 495, 525, 533, 547, 623
- Double-strand breaks, 545
- Dow AgroSciences, 587
- Drosophila melanogaster*. *See* Fruitfly
- Drought resistance, 211–212, 581
- Drought tolerance, 53, 211–212, 580–581
- Drought Tolerant Maize for Africa project, 52, 581
- E**
- Effective population size (Ne), 328, 330, 491, 494
- Effector-triggered immunity (ETI), 622
- Elaeis guineensis*. *See* Oil palm, African
- EM algorithm, 121–123
- Embryo
- Embryo culture, 508
  - Embryo sac, 30, 80, 88, 385
- Emerson, Rollins Adams, 33, 115
- Endosperm, 12, 80, 82, 85, 116–118, 207, 292, 372, 385, 501, 502, 587, 589
- Endosperm Balance Number (EBN), 501
- Epigenetic variation, maize inbreds, 158–159
- Epigenetic variation, methylated regions, 159
- Epistasis, 134, 161, 197, 288, 297, 316, 317, 319, 325, 326, 333, 390–392, 394, 397, 406, 427, 429, 433–435, 438, 439, 446, 447, 451, 460, 467–469, 471, 472, 477
- Escherichia coli*, 45, 100, 137, 139, 562, 564, 566, 580
- Essentially derived (cultivar) variety, 40
- European Food Safety Authority (EFSA), 569
- Eutrophication, 68
- Expression Quantitative Trait Loci (eQTLs), 193
- F**
- F<sub>1</sub> hybrids, 93
- Faba bean (*Vicia faba*), 275, 340, 497
- Fairchild, Thomas, 32
- FAO. *See* Food and Agricultural Organization (FAO)
- Farmers, Amerindian in Amazonia, 278
- Farmers, Benin, 280
- Farmers, Duupa in Cameroon, 274, 278
- Farmers, Incas in the Andes, 26
- Farmers, Quechua people, 489
- Farmers' Rights, 487
- Farming, Abu Hureyra, 5
- Farming, bringing land into cultivation, 5–8
- Farming, conventional, 219, 228–229
- Farming, fate of wild nature, 624–625
- Farming, four-course rotation, 26

- Farming, Great Plains of the USA, 19, 25, 67, 473  
 Farming, intensive, 6, 625, 626  
 Farming, land available, 54, 55  
 Farming, organic, 218, 219, 228–229, 444  
 Farming, systems, 5, 26, 208, 228, 373, 422, 444, 483, 484, 624–626  
 Farming, transition to, 4–5  
 Farming, wet fens, 6, 7  
 Fertilizers, chemical, 26  
 Fertilizers, nitrogen, 46, 49, 57, 61, 68, 69, 156, 217–219, 325  
 Fertilizers, phosphate, 69, 220  
 Fescue, meadow (*Festuca pratensis*), 186, 260  
*Festolium*, 260–261  
*Festuca pratensis*. *See* Fescue, meadow  
*Ficus thunbergii*, 544  
 FISH, 84–85, 259  
 Fisher, Ronald A., 40, 114, 120, 161, 164  
 Flanking sequence tags (FSTs), 553  
 Flax (*Linum usitatissimum*), 596  
 Flax, rust (*Melampsora lini*), 596–598  
 Flor, Harold H., 597  
 Flow cytometry, 257, 385, 416  
 Flowering plants, 29–30, 78  
 Flowering plants, flower structure, 30  
 Flowering time. *See* Heading date  
 Fluorescence in situ hybridization. *See* FISH  
 Fodder beet (*Beta vulgaris*), 27, 203, 496  
 Food and Agricultural Organization (FAO), 6, 42, 49, 54, 59, 60, 67, 373, 485, 487, 491, 530, 540  
 Food production, 61–62  
 Food production, factors affecting, 58–61  
 Food production, land area, 59–61  
 Food traders, 29  
 Ford Foundation, 48  
 Fortune, Robert, 19  
*Fragaria × ananassa*. *See* strawberry, pine  
*Fragaria chiloensis*, 377  
*Fragaria vesca*. *See* Strawberry, Wood  
*Fragaria virginiana*, 377  
*Fusarium oxysporum* f. sp. *conglutinans*. *See* Cabbage, Fusarium wilt  
 Fruit, aggregate, 83  
 Fruit, caryopsis, 83, 547  
 Fruit, dry, 83  
 Fruit, fleshy, 83  
 Fruit, multiple, 83  
 Fruit, simple, 82  
 Fruit, structure, 82  
*Fruitfly* (*Drosophila melanogaster*), 42, 116, 529  
 Fungi, ascomycete, 87, 89, 117, 509  
*Fusarium* head blight resistance. *See* Wheat  
*Fusarium oxysporum* f.sp. *cubense*. *See* Banana wilt or Panama disease
- G**  
 Gametophytes, 77–80, 542  
 Gates Foundation, Bill and Melinda, 54  
 Gaud, William, 47  
 Gauss, Carl Friedrich, 164  
 Gene, exons, 139, 144, 553  
 Gene, introns, 139, 155, 572, 574, 579  
 Gene, stacking, 567, 583  
 Gene expression  
   Gene expression, enhancers and silencers, 140  
   Gene expression, genome-wide analysis, 157–158  
   Gene expression, leucine zipper transcription factors, 140  
   Gene expression, promoter. *See* Promoter  
   Gene expression, RNA polymerase II, 139, 140  
   Gene expression, transcription factor, 139–141  
   Gene expression, transcription factor AP2, 145  
   Gene expression, transcription factor DOF, 150  
   Gene expression, transcription factor R2R3 MYB, 152  
   Gene expression, transcription factor WRKY, 144  
   Gene expression, transcription factor YABBY, 145, 155  
   Gene expression, transcription factor Zinc Finger, 140  
   Gene expression, transcriptomics, 157  
   Gene silencing  
     Gene silencing, post-transcriptional, 573, 574, 576  
     Gene silencing, RNA interference, 573  
 Genebanks  
   Genebanks, core collections, 493  
   Genebanks, Genesys, 492  
   Genebanks, size of core collections, 494  
   Genebanks, standards, 491, 492  
   Gene-pools, 32  
   General Mixed Model, 162, 169–174  
   General Mixed Model, Best Linear Unbiased Estimators (BLUEs), 171  
   General Mixed Model, Best Linear Unbiased Predictors (BLUPs), 171, 172  
 Genetic code, 135–137, 499  
 Genetic drift, 13, 274, 325, 327, 494  
 Genetic linkage. *See* Linkage

- Genetic modification. *See* Genetic transformation
- Genetic recombination, 89–91
- Genetic recombination, chromatid interference, 118
- Genetic recombination, coincidence, 118
- Genetic recombination, crossing-over, 102, 117
- Genetic recombination, cross-over position interference, 118
- Genetic recombination, crossovers, 117, 118
- Genetic recombination, Holliday junction, 91
- Genetic recombination, homologous recombination, 90
- Genetic recombination, molecular basis, 89–91
- Genetic recombination frequencies, 118, 119, 123, 128–130, 187, 447
- Genetic recombination frequencies, estimation, 120–123
- Genetic recombination frequencies, maximum likelihood estimate, 120, 130
- Genetic resources, 14, 32, 47, 484–487, 492, 495–498
- Genetic resources, Multilateral System (MLS), 487
- Genetic resources, Nagoya Protocol, 487
- Genetic resources, utilization, 487
- Genetic transformation, 45, 75, 93, 141, 149, 152, 213, 257, 360, 366, 370, 404, 484, 485, 498, 505, 561, 579, 581, 583, 587, 599
- Genetic transformation, *Agrobacterium*-mediated, 562, 563
- Genetic transformation, biolistic-mediated, 565
- Genetic transformation, clean-gene technology, 567
- Genetic transformation, techniques, 562
- Genetic variation, 21, 41, 75, 91, 97, 148, 158, 164, 165, 193, 262, 289, 309, 319, 326, 328, 329, 343, 345, 346, 351–353, 366, 368, 374, 395, 399, 425, 427, 436, 444, 449, 455, 485, 516
- Genetic variation, additive, 41, 326, 399, 436
- Genetic variation, between and within crosses, 427–431
- Genetic variation, non-additive, 41, 343, 345
- Genetic vulnerability, 485
- Genetically modified crops (GM)
- Genetically modified crops, *Bt*-crops, 577, 578
- Genetically modified crops, cisgenic, 572
- Genetically modified crops, cultivation in centre of origin and diversity, 585–586
- Genetically modified crops, environmental risk assessment (ERA), 569, 573
- Genetically modified crops, EU legislation, 568–569
- Genetically modified crops, food and feed risk assessment, 568
- Genetically modified crops, herbicide tolerance (HT), 574
- Genetically modified crops, insect-resistant (IR), 577–580
- Genetically modified crops, intragenic, 572
- Genetically modified crops, Monarch butterfly affair, 579
- Genetically modified crops, Pusztai affair, 579
- Genetically modified crops, regulations, 568
- Genetically modified crops, transgenic, 572
- Genetically modified crops, USA legislation, 568
- Genome evolution, 106, 107, 213–269
- Genome evolution, cereal, 261
- Genome evolution, copy number variants, 267
- Genome evolution, present/absent variants, 267, 268
- Genome wide association (mapping) studies (GWAS), 193, 197
- Genomic estimated breeding values (GEBVs), 194
- Genomic *in situ* hybridization (GISH), 84–85, 186, 187, 261, 265, 500, 506
- Genomic selection, 161–206, 255, 310, 366, 383–384
- Genomic selection, apple breeding, 384
- Genomic selection, cross-validation, 202–203
- Genomic selection, kinship, 201–202
- Genomic selection, linkage disequilibrium, 198
- Genomic selection, methods, 202
- Genomic selection, population structure, 198–202
- Genomic selection, practical issues, 206
- Genomic selection, results in practice, 203–205
- Genomic selection, tropical maize breeding, 205
- Genotype × environment interactions, AMMI model, 221
- Genotype × environment interactions, analysis, 209–214
- Genotype × environment interactions, concept, 209–210
- Genotype × environment interactions, correlated response to selection, 227–229

- Genotype × environment interactions, detection, 210–214
- Genotype × environment interactions, principal component analysis, 221–227
- Genotype × environment interactions, regression analysis, 214–221
- Genotype × environment interactions, response curves, 216–218
- Germplasm Enhancement of Maize (GEM) Project, 524–525
- Gilbert, Wally, 43, 77
- GISH. *See* Genomic *in situ* hybridization (GISH)
- Global Crop Diversity Trust (GCDT), 486–488
- Globodera pallida*. *See* Potato cyst nematode, white
- Globodera rostochiensis*. *See* Potato cyst nematode, golden
- Glycine max*. *See* Soybean
- Gossypium* spp.. *See* Cotton
- Grapevines (*Vitis vinifera*), 153, 259, 379, 381, 382, 497
- Grapevines, black rot (*Guignardia bidwellii*), 381
- Grapevines, breeding, 379–382
- Grapevines, downy mildew (*Plasmopara viticola*), 381
- Grapevines, hybrid direct producers, 381
- Grapevines, phylloxera root louse (*Daktulosphaira vitifoliae*), 381
- Grapevines, powdery mildew (*Uncinula necator*), 381
- Grapevines, red grapes, 153–154
- Grapevines, white grapes, 153–154
- Graphical genotyping, 177–178, 181
- Greenhouse gas emissions. *See* Climate change
- Green Revolution, 26, 45, 47–50, 164, 440, 445, 619
- Green Revolution genes, 155–157
- Green Revolution genes, dwarfing in wheat (*Rht-B1/Rht-D1*), 156
- Green Revolution genes, semidwarfing in rice (*sd-1*), 156
- Grew, Nehemiah, 29
- Guignardia bidwellii*. *See* Grapevine, black rot
- Gymnosperms, 79, 84
- H**
- Haemanthus katherinae*, 85
- Haplopappus gracilis*, 77
- Haplotypes, 148, 192, 193, 264, 267
- Hardy, Godfrey Harold, 284
- Heading date, 192, 229, 231, 309
- HEALTHGRAIN programme, 54
- Heat tolerance, 67
- Helianthus annuus*. *See* Sunflower
- Helianthus petiolaris*. *See* Sunflower, wild
- Helicoverpa armigera*. *See* Cotton, bollworm
- Helicoverpa armigera*, 578
- Herbicide tolerance, 52, 137, 138, 532, 547, 561, 567, 574, 575
- Herbicide tolerant, 52, 137–138, 532, 575–576
- Herbicides, 26, 46, 52, 70, 137, 138, 166, 456, 567, 574–576
- Herbicides, 2,4,5-T, 52, 70
- Herbicides, 2,4-D, 52, 70
- Herbicides, cyclohexanedione-tolerant, 137
- Herbicides, selective, 26, 576
- Herbicides, sulphonylurea-tolerant, 137
- Herbicides, triazine-tolerant, 137, 138, 227
- Heritability, 41, 149, 167, 168, 172, 186, 204, 205, 227–229, 255, 304, 306, 307, 312, 313, 318, 319, 321–323, 328, 350–354, 358, 363, 364, 369, 370, 374, 378–400, 430, 448–450, 453, 522, 605
- Heritability, broad sense, 167, 168, 229, 350, 358, 363
- Heritability, narrow sense, 344, 364, 449, 450, 605
- Heterosis, 0, 47, 51, 246, 248, 267, 310, 311, 341, 363, 387, 388, 392, 404, 406, 410, 421, 433, 460–463, 467, 469, 470, 472, 476
- Heterosis, better parent, 311, 363, 406, 421, 433, 460–463, 467, 470, 472, 476
- Heterosis, dominance hypothesis, 11, 13, 14, 16, 40, 88, 113, 114, 121, 184, 237, 430, 501, 596, 600
- Heterosis, in maize, 406, 409
- Heterosis, in rice, 138, 142, 156, 231, 266
- Heterosis, in tomato, 472
- Heterosis, midparent, 469
- Heterosis, overdominance hypothesis, 433, 462
- Heterosis, predicting, 41, 312, 406, 407
- Heterotic groups, 268, 401, 404, 406, 410, 411, 413, 414, 422, 470, 524, 526
- Hevea brasiliensis*. *See* Rubber
- Histogenic layers, 81
- Hofmeister, Wilhelm, 37, 77
- Holocene Era, 4
- Hooke, Robert, 36
- Hooker, Joseph, 20

- Hopkins, Cyril G, 323  
*Hordeum bulbosum*, 498  
*Hordeum vulgare*. *See* barley 146, 497, 498, 621  
*Hordeum vulgare* ssp. *spontaneum*, 146  
Hovey, Charles, 377  
Human diets, 27, 28  
Human nutrition, 68, 80  
Human under nutrition, 54  
Hunger, 50, 54, 257  
Hunter-gatherers, 3–5, 10, 16, 486  
Hybrid cultivars, 28, 35, 265, 267, 293, 311, 331, 387–423, 459–480, 523, 554  
Hybrid cultivars, double-cross, 46, 47, 307, 310, 387, 394, 414  
Hybrid cultivars, single-cross, 46, 47  
Hybrid cultivars, triple-cross, 414  
Hybrid vigour, 35, 47, 246, 267, 387, 410, 459 *See also* Heterosis  
Hybridization, deliberate, 33, 49, 378, 381, 383  
Hybridization, First International Conference, 33  
Hybridization, mechanics of, 33–34  
Hybridization, strategy, 344–346, 348
- I**  
Illinois Long Term Selection Experiment (ILTSE), 387  
Inbreeding, 35, 123, 187, 198, 258, 263, 276, 278, 285, 309, 312, 338, 357, 359, 377, 389, 390, 421, 444, 479  
Inbreeding depression, 47, 102, 390–393, 397, 418, 425, 459, 475, 505, 521  
Incorporation. *See* base broadening, 75, 88, 94, 265, 279, 484, 497, 505, 517, 520, 521, 526, 547  
Ingen-Housz, Jan, 63  
Inheritance, independent assortment, 114, 116, 133  
Inheritance, laws of, 34, 35, 109  
Inheritance, segregation, 40, 110, 114  
Initial cells, 533–535  
Insecticide resistance, 577–578  
Insecticides, 0, 52, 70, 371, 571, 577  
Insecticides, DDT, 70  
Insecticides, neonicotinoid, 70  
Intergovernmental Panel on Climate Change (IPCC), 61, 65  
Interior Spruce (*Picea engelmannii* × *glauca*), 203  
International Atomic Energy Agency (IAEA), 42, 530  
International Board for Plant Genetic Resources (IBPGR), 47, 485  
International Centre for Agricultural Research in the Dry Areas (ICARDA), 526  
International Centre of Tropical Agriculture (CIAT), 369  
International Energy Agency (IEA), 61  
International Institute for Applied Systems Analysis (IIASA), 6, 60  
International Institute of Tropical Agriculture (IITA), 369, 581  
International Maize and Wheat Improvement Centre, Mexico (CIMMYT), 48, 50, 203, 206, 207, 212, 219, 440, 445, 581, 619, 620  
International Potato Centre (CIP), 484, 494  
International Research Centres, 45  
International Rice Research Institute (IRRI), 49, 440, 590  
International Seed Federation (ISF), 36  
International Seed Testing Association (ISTA), 36  
International Treaty on Plant Genetic Resources, 486–488  
Introgression, 11, 71, 186, 187, 189, 221, 361, 380, 472, 483–527, 573, 583, 584, 594, 600, 605  
Introgression, lettuce, 507  
Introgression, onion, 506–507  
Introgression, potato, 18  
Introgression, sugar beet, 30  
Introgression, tomato, 508–509  
*Ipomoea batatas*, 497 *See also* Sweet potato  
Ishiwatari, Shigetane, 577  
Isolation distances, 36
- J**  
Johannsen, Wilhelm, 115  
JoinMap, 127, 128, 130  
J.R. Simplot Company, 584  
Jukes-Cantor model. *See* molecular evolution
- K**  
Kale (*Brassica oleracea*), 41  
Kale breeding, 329  
Kale, population improvement, 307  
Kale, triple-cross hybrids, 414  
Kentucky bluegrass (*Poa pratensis*), 385

Knight, Thomas Andrew, 33, 360  
 Kölreuter, Joseph, 32

## L

*Lactuca saligna*, 507, 508  
*Lactuca sativa*. See lettuce, 497  
*Lactuca serriola*, 507  
*Lactuca virosa*, 507, 508  
 Landraces  
 Landraces, cross-pollination, 274–280  
 Landraces, genetic structure, 273–290  
 Landraces, metapopulations, 274  
*Lathyrus odoratus*. See Sweet peas  
 Latin American Maize Project  
 (LAMP), 524  
 Lavoisier, Antoine, 63  
*Leptinotarsa decemlineata*, 582  
 Lettuce (*Lactuca sativa*), 497, 507  
 Lettuce, downy mildew (*Bremia lactucae*),  
     381, 506, 507  
 Lettuce breeding, 507, 508  
 Life cycle of plants, 75  
 Life cycle of plants, angiosperms, 79  
*Lilium longiflorum*, 88  
 Limagrain, 397  
 Linkage analysis, 123, 249, 253  
 Linkage analysis, ordered tetrad, 117  
 Linkage analysis, tetrasomic inheritance,  
     248, 253  
 Linkage disequilibrium, 193, 194, 198,  
     200–202, 288, 312–316, 467,  
     494, 525, 526  
 Linkage disequilibrium, amount, 200  
 Linkage disequilibrium, mixed selfing and  
     random mating, 289  
 Linkage disequilibrium, population  
     improvement, 312–316  
 Linkage group, 101, 116, 117, 127–130,  
     185, 187, 253, 426, 617  
 Linkage group, ordering loci, 127–130  
 Linkage maps, 106, 109–132  
 Linkage maps, BINS, 131  
 Linkage maps, Haldane mapping function,  
     124, 126  
 Linkage maps, high density, 130–132  
 Linkage maps, Kosambi mapping  
     function, 126  
 Linkage maps, mapping functions,  
     118, 124  
 Linkage, partial, 116  
 Linnaeus, Carl, 30  
*Linum usitatissimum*, 596 See also Flax

*Lolium multiflorum*, 260, 261 See also  
 Ryegrass, Italian  
*Lolium perenne*, 186, 192, 308 See also  
 Ryegrass, perennial

## M

*Magnaporthe grisea*, 456  
*Magnaporthe oryzae*, 509  
 Maize (*Zea mays*), 12, 17, 116,  
     143, 497, 522  
 Maize, copy number variants, 267  
 Maize, cytoplasm C, 403, 404  
 Maize, cytoplasm S, 403, 404  
 Maize, cytoplasm T, 402  
 Maize, cytoplasmic male sterility, 404  
 Maize, double-cross hybrids, 46, 307, 394  
 Maize, drought tolerant, 580  
 Maize, European corn borer (*Ostrinia*  
     *nubilalis*), 579  
 Maize, F<sub>1</sub> hybrid breeding, 469  
 Maize, gene pools, 498–499  
 Maize, genome duplication, 266  
 Maize, genome evolution, 233–269  
 Maize, germplasm enhancement, 522–526  
 Maize, herbicide tolerant, 52  
 Maize, heterosis, 469, 471, 472, 476  
 Maize, heterotic groups, 406, 409–410  
 Maize, hybrid, 410, 413  
 Maize, hybrid seed production, 415  
 Maize, inbred lines, 479, 508  
 Maize, insect resistant, 579–580  
 Maize, long-term selection, 323, 329  
 Maize, pan-genome, 267, 269  
 Maize, pan-transcriptome, 269  
 Maize, present/absent variants  
 Maize, seed production, 531  
 Maize, single-cross hybrids, 394, 397  
 Maize, SmartStax™, 567, 577  
 Maize, southern corn leaf blight  
     (*Bipolaris maydis*), 402, 404  
 Maize, subgenomes, 268  
 Maize, teosinte (*Zea mays* subspecies  
     *parviflora*), 143, 522  
 Maize, transposable elements, 158, 266  
 Maize, western corn rootworm (*Diabrotica*  
     *virgifera virgifera*), 577  
 Maize, YieldGard, 579  
 Malthus, Thomas, 58  
*Malus × domestica*. See Apple  
*Malus sieversii*, 383  
 Mandarins (*Citrus clementina* x  
     *C. tangerina*), 257

- Manihot esculenta*. See Cassava
- Markham, Clement, 20
- MASIPAC network, 444
- Mating design, 309, 463, 468
- Mating design, diallel, 176, 302
- Mating design, North Carolina Experiment III (NCIII), 463, 464
- Mating design, Triple Test Cross (TTC), 468, 470
- Mating systems, 34, 35, 39, 273, 280, 281, 284, 296, 341, 492
- Maturity, 25, 149, 150, 205, 211, 216, 255, 292, 380, 399, 415, 445, 458, 473, 521, 523, 531, 547, 566, 606, 608, 617
- Maturity, potato, 496, 497, 503, 504
- Maturity, potato gene (*StCDF1*), 152
- Max Plank Institute, 602
- Maximum likelihood, estimation, 119, 171
- Maximum likelihood, common logarithm of likelihood ratio (LOD), 121, 127, 184
- Maximum likelihood, likelihood ratio (odds), 184
- Maximum likelihood, restricted (REML), 202, 203, 211
- McClintock, Barbara, 106
- Medicago sativa*. See Alfalfa
- Megaspores, 78–80, 243
- Meiosis, 38, 39, 75, 78–80, 84, 87–89, 91–93, 101, 103–105, 111, 117, 118, 123, 190, 241–244, 248–250, 262, 286, 301, 385, 443, 510, 517, 519, 547, 597
- Meiosis, chromatid exchange, 89
- Meiosis, coefficient of double reduction, 249
- Meiosis, double reduction, 253
- Meiosis, Janssens-Belling theory, 89
- Meiosis, quadrivalents, 104, 248, 249
- Melampsora lini*, 596
- Mendel, Gregor, 34
- Mendel, laws of inheritance. See Inheritance
- Mendel's genes (*A, I, Le* and *R*), 141–142
- Meristem, 81, 82, 86, 96, 150, 367, 370, 371, 546, 551
- Meristem, apical, 551
- Meristem, initials, 96
- Meristem-tip culture, 367
- Metabolomics, 157
- Methane, 58, 59, 61, 530
- Microarrays, 157–158, 193
- Microscopes, 36, 84
- Microscopes, electron, 37
- Microscopes, optical, 37
- Mimulus guttatus*, 101 See also Yellow monkeyflower
- Mitochondrion, 37, 233, 240, 241, 404, 416, 420, 477, 500, 531
- Mitochondrion genome, 233, 416, 420
- Mitosis, 38, 78, 80, 84–88, 117, 241
- Mixtures, 84, 85, 274–289, 425–458
- Mixtures, cultivar or variety, 11
- Mixtures, spring barley, 281, 439, 621
- Molecular clock, 233, 235, 237, 238, 240
- Molecular evolution, 233, 236, 240
- Molecular evolution, Jukes-Cantor model, 235–236, 238
- Molecular evolution, Markov model of base substitutions, 234–235
- Molecular marker, 13, 41, 45, 97, 148, 161, 177, 182, 183, 188, 190, 204, 255–257, 278, 294, 309, 366, 368, 385, 407, 422, 430, 469, 471, 507–512, 515, 552, 595, 604
- Molecular marker, AFLP, 421, 430, 506, 509
- Molecular marker, DARt, 100
- Molecular marker, RAPD, 98–100, 186, 506
- Molecular marker, RFLP, 98–100, 106
- Molecular marker, SCAR, 99
- Molecular marker, SNP, 101, 130, 132, 203, 254
- Molecular marker, SSR, 614
- Molecular marker, STS, 99
- Monsanto, 52, 525, 567, 575, 579, 580, 582
- Morgan (M), definition, 116
- Morgan, Thomas Hunt, 116
- Muller, Hermann Joseph, 529
- Multilateral System, 486
- See Genetic resources
- Multilines, 599
- Multiple-Interval-Mapping (MIM), 471
- Munerati, Ottavio, 505
- Musa acuminate*, 497
- Musa balbisiana*, 497
- Musa spp.*, 255
- Mustard, black (*Brassica nigra*), 258
- Mustard, Ethiopian (*Brassica carinata*), 258
- Mustard, Indian or brown (*Brassica juncea*), 258
- Mutagenesis, 42, 71, 81, 82, 87, 531–535, 539, 540, 542–556, 575, 622
- Mutagenesis, chemical, 533, 539, 545, 546, 548, 550, 552
- Mutagenesis, gamma ray, 42, 530, 540
- Mutagenesis, heavy-ion beam, 540, 543–544
- Mutagenesis, insertional, 553
- Mutagenesis, organelle genomes, 531
- Mutagens, 42, 530–533, 544–548
- Mutagens, alkylating agents, 545

- Mutagens, chemical, 545–548  
 Mutagens, ethyl methane sulphonate (EMS), 42, 530, 545  
 Mutagens, 1-ethyl-1-nitrosourea (ENU), 545  
 Mutagens, gamma-ray, 539–542  
 Mutagens, ionizing electromagnetic radiation, 540  
 Mutagens, ionizing particles, 540  
 Mutagens, 1-methyl-1-nitrosourea (MNU), 545  
 Mutagens, mustard gas, 42  
 Mutagens, physical, 42  
 Mutagens, sodium azide, 545, 546  
 Mutagens, X-rays, 42, 530  
 Mutants, gamma-ray, 42, 530, 539–543  
 Mutants, heavy-ion beam, 544–545  
 Mutants, identification, 82  
 Mutants, identification by TILLING, 548–550  
 Mutants, radiation-induced, 530  
 Mutation breeding, 539, 540, 542, 543, 559  
 Mutation breeding, MutMap, 552  
 Mutation breeding, population sizes, 14, 516, 536  
 Mutation experiments, 531–539  
 Mutation experiments, design, 189  
 Mutation rates, 534, 536, 538, 552  
 Mutation rates, estimation, 120  
 Mutations  
   Mutations, base-pair substitutions (SNPs), 97  
   Mutations, frame-shift, 150  
   Mutations, insertions and deletions (InDels), 97, 98, 559  
   Mutations, inversions, 104, 106  
   Mutations, translocations, 94, 97, 101, 104, 106, 265, 518, 557  
 Mycorrhiza, arbuscular, 69, 373
- N**  
 Nagoya Protocol. *See* Genetic resources  
 National Agricultural Research Systems (NARS), 51, 520, 522  
 The Natufians, 5  
 Nearest neighbour (spatial) analysis, 127, 168 *See also* Yield trials  
 Near-infrared reflectance spectroscopy (NIRS), 176  
 Neolithic Revolution, 3  
*Neurospora crassa*, 89, 117  
*Nicotiana alata*, 499  
*Nicotiana benthamiana*, 623  
*Nicotiana debneyi*, 499  
*Nicotiana forgetiana*, 499  
*Nicotiana glauca*, 499  
*Nicotiana langsdorffii*, 499  
*Nicotiana longiflora*, 40  
*Nicotiana rustica*, 216, 438, 452  
*Nicotiana tabacum*. *See* Tobacco  
*Nicotiana × sanderae*, 499  
 Nitrate Vulnerable Zones, 69  
 Nitrogen fixation, 68–69, 208, 311  
 Nitrogen use efficiency (NUE), 218–219  
 Nitrogen-fixing bacteria, 69  
 Nitrous oxide, 59, 61  
 Non-allelic interactions, 134, 288, 318, 390, 461, 462, 468  
 Nordic Genetic Resource Centre (NordGen), 488  
 Normal distribution, 164, 165, 354, 355, 433  
 North Carolina Experiment III (NCIII). *See* Mating design  
 Nucleases, 555, 557, 558, 560  
 Nucleases, mega- (MNs), 555  
 Nucleases, site-directed (SDNs), 555  
 Nucleases, transcription activator like effector (TALENs), 555, 557  
 Nucleases, zinc finger (ZFNs), 555
- O**  
 Obesity, 27, 54  
 Ohalo II Stone Age site, 4  
*Oidium neolyccopersici*. *See* tomato, powdery mildew  
 Oil palm, African (*Elaeis guineensis*), 20  
 Oilseed rape (*Brassica napus*), 27, 255, 259  
 Oilseed rape, CMS systems, 477  
 Oilseed rape, herbicide tolerant, 137–138, 575  
 Oilseed rape, HOLL types, 28  
 Oilseed rape, hybrids, 32, 47, 51, 479  
*Olea europaea*, 375  
 Olive (*Olea europaea*), 6, 83, 375, 376  
 Olive breeding, 375, 376  
 Onions (*Allium cepa*), 208, 417  
 Onion breeding, 417  
 Onions, cytoplasmic male sterility (CMS-S), 416–418  
 Onions, downy mildew (*Peronospora destructor*), 506, 507  
 Onions, F<sub>1</sub> hybrids, 387, 477  
*Ophioglossum petiolatum*, 77  
 Ordered tetrad analysis. *See* Linkage  
 Organic farming. *See* Farming  
*Oryza glaberrima*, 11 *See* rice, African  
*Oryza rufipogon*, 147, 190 *See* rice, wild

- Oryza sativa*, 21, 497 *See* rice, Asian  
*Ostrinia nubilalis*, 579  
Ovary, 30, 80, 82  
Ovules, 30, 79, 80  
Ozone, 65
- P**
- Pan-genome concept, 266–268  
*Papaver dubium*, 287  
Papaya (*Carica papaya*), 576  
Papaya ringspot virus (PRSV) resistance, 576  
Participatory plant breeding (PPB), 53, 229, 439, 445, 526–527  
Pathogens, 70, 174, 370, 381, 456, 570, 592, 599, 606, 621, 622, 624  
Pathogens, Avr3a effector of *P. infestans*, 599  
Pathogens, effectors, 599  
Pathogens, physiological races, 595, 603  
Pea, garden (*Pisum sativum*), 34, 109, 497  
Peach (*Prunus persica*), 376  
Peanut, 187, 189–190, 485, 486, 489, 496  
Peanut, chromosome segment substitution lines, 187  
Peanut, cultivated (*Arachis hypogaea*), 189–190  
Peanut, wild (*Arachis duranensis*), 189  
Peanut, wild (*Arachis ipaënsis*), 189  
Pearson, Karl, 113  
Pedigree inbreeding with selection, 426, 440–441, 447, 448, 454, 455, 463, 474  
Pepper (*Capsicum annuum*), 43, 496, 623  
*Peronospora destructor*. *See* Onion, downy mildew  
*Peronospora tabacina*. *See* Tobacco, blue mould  
Permutation tests, 184–185  
Pest resistance, 49, 362, 445, 483, 484, 493, 496, 524, 561, 579, 595, 606, 624  
Pesticides. *See* Insecticides  
Phase-contrast photography, 85  
*Phaseolus vulgaris*. *See* Bean, common  
Phenotyping, High-throughput, 176–177  
Philippine Rice Research Institute (PhilRice), 590  
*Phoma foveata*. *See* Potato, gangrene  
Phosphorus use efficiency (PUE), 218, 220–221  
Photoperiod, 26, 49, 148–149, 230–232  
Photoperiod, insensitivity, 49, 148, 230  
Photoperiod, maize, 148–149, 230–231  
Photoperiod, sensitivity, 26, 148, 230–232  
Photoperiod, wheat, 148
- Photoperiod genes, maize (*ZmCCT*), 149  
Photoperiod genes, wheat (*Ppd-D1 and Ppd-B1*), 148  
Photorespiration, 63, 241  
Photosynthesis, 37, 50, 59, 62–64, 138, 219, 241, 372, 575  
Photosynthesis, C<sub>3</sub> pathway, 63  
Photosynthesis, C<sub>4</sub> pathway, 63  
Photosynthesis, crassulacean acid metabolism (CAM), 64  
Phylogenetic distance, 237  
Phylogenetic distance, Jukes-Cantor, 237–238  
Phylogenetic trees, 233–240  
Phylogenetic trees, construction, 105, 233, 238  
*Phytophthora infestans*. *See* Potato, late blight  
*Picea engelmannii* × *glauca*. *See* Interior spruce  
Pioneer Hi-Bred, 394, 397  
*Pisum sativum*. *See* Pea, garden  
Plantation crops, 20  
Plantation crops, tropical, 8, 19–20, 343, 372  
Plant & Food Research (PFR), 384  
Plant Breeders' Rights, 40, 330, 362  
Plant breeding, future goals, 53–54  
Plant classification, 31  
Plant classification, Angiosperm Phylogeny Group III, 31  
Plant classification, cladistics, 31  
Plant classification, phylogenetic, 31  
Plant classification, taxonomic, 32  
Plant nomenclature, 32  
Plant nomenclature, Latin names, 32  
Plasmids, 45, 562–564, 566, 576, 577, 585, 586  
*Plasmodiophora brassicae*. *See* Brassicas, clubroot  
*Plasmopara viticola*. *See* Grapevines, downy mildew  
Pleiotropy, 327, 454  
Pliny the Elder, 153, 378  
Plot interference, 174–176  
*Poa pratensis*. *See* Kentucky bluegrass  
*Podosphaera leucotricha*. *See* Apple, powdery mildew  
Pollen grains, 79, 80, 84, 416  
Pollination  
Pollination, cross, 10, 33, 35, 39, 71, 273, 275–280, 290, 294, 312, 341, 359, 360, 374, 385, 414, 418, 419, 426, 442, 473, 478, 493  
Pollination, insect, 277, 368, 476, 479  
Pollination, mentor, 502

- Pollination, mixture of self- and cross-, 273, 275–278, 290  
 Pollination, outcrossing rates, 274, 275, 277, 278  
 Pollination, wind, 35  
 Polymerase chain reaction (PCR), 77, 98–100, 187, 192, 416, 541, 544, 550, 567, 583, 586  
 Polyploidization, 41, 242, 243, 246, 259, 262, 264  
 Polyploids, 41  
 Polyploids, allopolyploids, 41, 85, 241, 242, 255, 263, 500, 517, 548, 550  
 Polyploids, amphidiploids, 242, 260, 263, 508  
 Polyploids, autopolyploids, 241, 344, 500  
 Polyploids, autotetraploids, 127, 242, 248–255, 279, 296, 318, 338, 344, 504  
 Polyploids, classification, 241–242  
 Polyploids, segmental allopolyploids, 242  
 Population improvement, 291–341  
 Population improvement, alfalfa, 295, 296, 301, 308–310, 338  
 Population improvement, epistasis, 297, 316, 319, 325, 326, 333  
 Population improvement, foundation population, 293, 328, 329  
 Population improvement, full-sib family selection, 296–307  
 Population improvement, half-sib family selection, 294, 296–307, 312–314, 320–322  
 Population improvement, kale, 293, 306, 307, 320–322, 329, 330  
 Population improvement, maize, 291–295, 306, 307, 313, 319, 323  
 Population improvement, mass selection, 294–307, 309, 310, 312, 313, 316, 317  
 Population improvement, mixed selfing and random mating, 312  
 Population improvement, perennial ryegrass, 308  
 Population improvement, red clover, 296, 310, 338  
 Population improvement, S<sub>1</sub> family selection, 296–307  
 Population improvement, white clover, 311  
 Population structure, 192, 198, 200–202  
 Population structure, Hardy-Weinberg equilibrium, 201  
 Population structure, mixed selfing and random mating, 281–285  
 Population structure, two loci, 285–287  
 Potato  
 Potato, acrylamide, 585  
 Potato, *Agrobacterium*-mediated transformation, 564–565  
 Potato, Amflora GM potato, 582  
 Potato, asparagine, 143, 585  
 Potato, black spot bruise, 584, 585  
 Potato, CIP populations B1 to B3, 522  
 Potato, Colorado potato beetle (CPB) (*Leptinotarsa decemlineata*), 582  
 Potato, cultivated (*Solanum tuberosum*), 13, 243  
 Potato, Fortuna GM potato, 583  
 Potato, gangrene (*Phoma foveata*), 608  
 Potato, genetically modified, 582–586  
 Potato, germplasm collections, 484, 485  
 Potato, Group Andigena, 18, 494  
 Potato, Group Phureja, 361  
 Potato, Group Stenotomum, 13, 242  
 Potato, Group Tuberousum, 494, 503  
 Potato, Innate GM potato, 584  
 Potato, late blight (*Phytophthora infestans*), 361, 582–584, 599, 600  
 Potato, minitubers, 367  
 Potato, Neotuberosum, 521  
 Potato, NewLeaf™, 582  
 Potato, powdery scab (*Spongospora subterranea*), 608  
 Potato, reducing sugars, 585  
 Potato, seed certification, 366  
 Potato, seed potato production, 366–368  
 Potato, selection during clonal generations, 368, 369, 374, 375  
 Potato, single-node cuttings, 367  
 Potato, skin colour, 23  
 Potato, skin colour genes (D, R, and P), 152  
 Potato, true potato seed (TPS), 360  
 Potato, wart (*Synchytrium endobioticum*), 602  
 Potato, wild, 500, 503, 600  
 Potato, wild (*acaule*), 503  
 Potato, wild (*albicans*), 500  
 Potato, wild (*bukasovii*), 13, 22  
 Potato, wild (*bulbocastanum*), 503, 504, 583, 602, 616  
 Potato, wild (*chacoense*), 503  
 Potato, wild (*commersonii*), 502, 504  
 Potato, wild (*demissum*), 361, 500, 503, 504, 583, 600, 602, 616  
 Potato, wild (*hougasii*), 500  
 Potato, wild (*microdontum*), 361  
 Potato, wild (*spegazzini*), 503  
 Potato, wild (*stoloniferum*), 501, 583  
 Potato, wild (*venturii*), 583

- Potato, wild (*vernei*), 361, 503, 504  
Potato, wild (*verrucosum*), 502, 585  
Potato breeding, 243, 248, 260–368,  
    371, 374, 484, 503, 521, 602, 614  
Potato breeding, breeding scheme,  
    362–365, 371  
Potato breeding, crossing strategy,  
    344, 363, 364  
Potato breeding, early generation selection,  
    364, 365  
Potato cyst nematode, 183, 255, 364,  
    602, 606  
Potato cyst nematode, golden (*Globodera*  
    *rostochiensis*), 361, 594, 602  
Potato cyst nematode, white (*Globodera*  
    *pallida*), 361, 364, 594, 602  
*Potato leafroll virus (PLRV)*, 205, 582, 594  
*Potato virus X (PVX)*, 205, 602  
*Potato virus Y (PVY)*, 205, 582, 602, 623  
Priestly, Joseph, 63  
Promoters, 139–140, 149, 151, 154, 193,  
    553, 554, 556, 557, 564, 572, 574,  
    575, 579, 583  
Promoter, cauliflower mosaic virus 35S  
    (CaMV 35S), 150, 564, 582  
Proteomics, 157  
Protoplast fusion, 416, 417, 477, 498–500,  
    503, 573  
*Prunus persica*. *See* peach  
*Puccinia graminis*. *See* Wheat, stem rust  
*Puccinia striiformis*. *See* Wheat, yellow rust  
*Puccinia triticina*. *See* Wheat, leaf rust  
Pure line selection, 49, 71, 425–426
- Q**  
Quantitative Trait Locus (QTL), 40–41, 102,  
    123, 143, 144, 147, 149, 161, 177–181,  
    183–187, 189–195, 198, 199, 202, 203,  
    206, 213, 231, 232, 253–254, 325, 366,  
    469–472, 508, 509, 516, 616, 623  
QTL analysis, backcross population, 177–181  
QTL analysis, chromosome segment  
    substitution lines, 187–190  
QTL analysis, composite interval mapping,  
    185–186  
QTL analysis, interval mapping, 179–182, 254  
QTL analysis, introgression mapping, 186  
QTL analysis, tetrasomic inheritance, 123  
QTL analysis, two linked QTLs, 181–183, 186  
QTL analysis, whole genome resequencing, 191  
Quantitative traits, 40, 41, 161, 163–166, 176,  
    177, 206, 306, 344, 364, 427, 431  
*Quinoa (Chenopodium quinoa)*, 54
- R**  
Radish (*Raphanus sativus*), 259  
Ramularia collo-cygni. *See* barley, Ramularia  
    leaf spot  
*Raphanobrassica*, 41, 258  
*Raphanus sativus*. *See* Radish  
Raspberry, 378  
Raspberry, breeding, 378, 379  
Raspberry, floricanes, 378  
Raspberry, primocanes, 378, 379  
Raspberry, red (*Rubus idaeus*), 378  
Ray, John, 31  
Reactive oxygen species (ROS), 214, 541  
Reciprocal recurrent selection, 411–413  
Reid, Robert, 292  
Response to selection. *See* selection response  
Reverse breeding, 93  
*Rhizobium leguminosarum* biovar *trifolii*, 310  
Rice  
    African (*Oryza glaberrima*), 12, 509  
    Asian (*Oryza sativa*), 11, 143, 497, 509  
    rice, bacterial blight (*Xanthomonas*  
        *oryzae*), 557  
    rice, basmati, 190, 426, 474  
    rice, biostatic-mediated transformation,  
        563, 566  
    rice, blast disease (*Magnaporthe grisea*,  
        *M. oryzae*), 191, 509  
    rice, breeding, participatory, 444  
    rice, chromosome segment substitution  
        lines, 190  
    rice, Golden, 52, 567, 587–590  
    rice, Golden Rice 2 (GR2), 588, 590  
    rice, heterosis, 459, 469–472, 474–475  
    rice, hybrids, 49, 50, 220, 456, 470, 474  
    rice, *Indica*, 11, 12, 144, 198, 231, 454,  
        470, 474, 588  
    rice, introgression, 497, 509–510  
    rice, *japonica*, 11, 12, 144, 231, 539, 541, 588  
    rice, new plant type, 49  
    rice, New Rice for Africa (NERICA),  
        509–510  
    rice, semi-dwarf, 49, 157, 426, 440  
    rice, short-duration cultivars, 26, 27  
    rice, wild (*Oryza rufipogon*), 190  
RIKEN ring cyclotron, 543  
RNA, artificial micro-RNA, 554  
RNA, interference, 141, 151, 553, 554,  
    572, 584–586  
RNA, messenger, 136  
RNA, pre-mRNA, 139  
RNA, ribosomal, 137, 265  
RNA, small, 140, 141  
RNA, small interfering, 573

- RNA, transfer, 135, 137  
 Rockefeller Foundation, 48, 49  
*Romalea microptera*, 88  
 Rubber (*Hevea brasiliensis*), 20  
*Rubus idaeus*. *See* raspberry, red  
 Rye (*Secale cereale*), 5, 263, 421–422  
 Rye, cytoplasmic male sterility, 422  
 Rye, hybrid, 59, 422  
 Rye, Pampa CMS, 423  
 Ryegrass, candidate genes, 260  
 Ryegrass, Italian (*Lolium multiflorum*), 260, 310  
 Ryegrass, perennial (*Lolium perenne*), 260, 308, 311
- S**
- Saccharomyces cerevisiae*, 89  
*Saccharum barberi*. *See* Sugar cane, Indian  
*Saccharum officinarum*. *See* Sugar cane  
*Saccharum robustum*. *See* Sugar cane, wild  
*Saccharum sinense*. *See* Sugar cane, Chinese  
*Saccharum spontaneum*. *See* Sugar cane, wild  
 Salt tolerance, 213–214, 496  
*Salvia horminum*, 133  
 Sanger, Fred, 43, 77  
 Savory, Allan, 60  
*Schizophyllum commune*, 216  
 Schleiden, Mathias, 36  
 Schwann, Theodor, 36  
 Scottish Crop Research Institute (SCRI), 167, 360, 363–365, 442, 446, 465  
 Scottish Plant Breeding Station (SPBS), 360, 364, 504, 600–602  
*Secale cereal*. *See* Rye  
 Seed  
 Seed, basic, 330, 367, 441, 478  
 Seed, certification, 36  
 Seed, certified, 36, 305, 311, 330, 332, 367, 414, 422, 478  
 Seed companies, 29, 394, 416, 417, 590  
 Seed, purity, 35–36  
 Seed, quality, 35–36, 308  
 Seed, structure, 37  
 Seed, trade, 29  
 Selection  
 Selection, altered chemical composition, 27–28, 422  
 Selection, colour, 152–155  
 Selection, correlated responses, 454  
 Selection, differential, 228, 299, 300, 303, 305, 318, 321, 350, 351, 354, 399, 448  
 Selection differential, standardized, 299, 300, 303, 350

- Selection, during domestication, 21, 565  
 Selection, during pedigree inbreeding, 446  
 Selection, effect on variance, 449–451  
 Selection, full-sib family, 329  
 Selection, half-sib family, 329  
 Selection, in new environments, 24–25  
 Selection, independent culling levels, 319, 354, 355, 357  
 Selection indices, 320  
 Selection indices, kale, 320  
 Selection indices, optimum index of Smith, 320  
 Selection, intensity of, 227, 228, 299, 300, 303, 313, 328, 345, 350, 354, 355, 358, 400, 430, 439, 449, 450  
 Selection limit, 325–328  
 Selection, long-term, 293, 313, 323, 326, 327, 329, 330  
 Selection, marker-assisted, 132, 183–184, 190, 194, 203, 206, 212, 213, 366, 370, 507, 509  
 Selection, mass, 295  
 Selection, maximum response to, 328  
 Selection, multistage, 343–346, 348, 350, 351, 353–355, 357–387, 454  
 Selection, multi-trait, 354–358, 387  
 Selection, S1 family, 296–306  
 Selection, shape, 152–155  
 Selection, tandem, 319, 354–357  
 Selection, visual, 22–24, 219, 362, 364, 542  
 Selection response, 318  
 Selection response, equilibrium, 318  
 Selection response, immediate, 318  
 Self-incompatibility, 35, 258, 277, 279, 295, 310, 311, 331, 359, 414, 415, 422, 478, 500  
 Self-incompatibility, gametophytic, 279, 310, 311, 331, 359, 383, 422, 500  
 Self-incompatibility, sporophytic, 258, 295, 359, 478  
 Shirreff, Patrick, 33  
 Single seed descent (SSD), 123, 400, 426, 440–449, 451–455, 463, 470, 476, 479, 507  
*Sinorhizobium meliloti*, 309  
 Solanaceae Genome Network (SGN), 131  
*Solanum acaule*. *See* Potato, wild  
*Solanum albicans*. *See* Potato, wild  
*Solanum brevidens*, 504  
*Solanum bukasovii*. *See* Potato, wild  
*Solanum bulbocastanum*. *See* Potato, wild  
*Solanum chacoense*. *See* Potato, wild  
*Solanum chilense*. *See* Potato, wild  
*Solanum commersonii*. *See* Potato, wild  
*Solanum demissum*. *See* Potato, wild

- Solanum habrochaites*. *See* Tomato, wild  
*Solanum hougasii*, 500, 501 *See* Potato, wild  
*Solanum lycopersicum*. *See* Tomato  
*Solanum microdontum*. *See* Tomato, wild  
*Solanum neorikii*. *See* Tomato, wild  
*Solanum pennellii*. *See* Tomato, wild  
*Solanum peruvianum*. *See* Tomato, wild  
*Solanum pimpinellifolium*. *See* Tomato, wild  
*Solanum spegazzini*. *See* Potato, wild  
*Solanum stoloniferum*. *See* Potato, wild  
*Solanum tuberosum*. *See* Potato, cultivated  
*Solanum venturii*. *See* Potato, wild  
*Solanum vernei*. *See* Potato, wild  
*Solanum verrucosum*. *See* Potato, wild  
Somaclonal variation, 82, 256, 366, 559  
Somatic fusion. *See* Protoplast fusion  
Somatic hybrids, 499, 503, 504  
*Sordaria fimicola*, 89, 117  
Sorghum (*Sorghum bicolor*), 144, 145, 278, 473, 497  
Sorghum, hybrid, 474  
Sorghum, wild (*Sorghum propinquum*), 144, 145  
*Sorghum bicolor*. *See* Sorghum  
*Sorghum propinquum*. *See* Sorghum, wild  
Soybean (*Glycine max*), 25, 496  
Soybean, breeding, 208  
Soybean, herbicide tolerant (HT), 575  
Soybean, insect resistant, 52, 577  
Species concept, 32  
Species concept, biological, 32  
Species concept, evolutionary, 32  
Species concept, taxonomic, 32  
*Spongospora subterranea*. *See* Potato, powdery scab  
Sporophyte, 77–80  
Stadler, Lewis John, 42, 118, 529  
Strasburger, Eduard, 80  
Strawberry, breeding, 377, 378  
Strawberry, pine (*Fragaria × ananassa*), 377  
Strawberry, wood (*Fragaria vesca*), 377  
Sturtevant, Alfred, 116  
Sugar beet, breeding, 423, 424  
Sugar beet, cytoplasmic-genetic male sterility, 422  
Sugar beet, diploid hybrids, 423  
Sugar beet, herbicide tolerant, 174, 575  
Sugar beet, hybrids, 423  
Sugar beet, leaf spot (*Cercospora beticola*), 505  
Sugar beet, monogerm, 422, 423  
Sugar beet, sugar content, 27  
Sugar beet, three-way cross, 423  
Sugar beet, triploid hybrids, 422, 423  
Sugar cane (*Saccharum officinarum*), 19, 264, 265, 372, 374, 497  
Sugar cane, biofuel, 59  
Sugar cane, breeding, 374  
Sugar cane, Chinese (*Saccharum sinense*), 264, 265  
Sugar cane, family selection, 375  
Sugar cane, Indian (*Saccharum barberi*), 264, 265  
Sugar cane mosaic virus (SCMV), 265  
Sugar cane, wild (*Saccharum robustum*), 264, 265, 374  
Sugar cane, wild (*Saccharum spontaneum*), 264, 265  
Sunflower (*Helianthus annuus*), 13, 18, 28, 478–480, 496  
Sunflower, high-oleic cultivars, 28  
Sunflower, hybrids, 478, 479  
Sunflower, wild (*Helianthus petiolaris*), 478, 479  
Sutton, Walter, 29, 38, 111  
Svalbard Global Seed Vault, 9, 488  
Swaminathan, Monkombu Sambasivan, 48, 249  
Swede (*Brassica napus*), 27, 174, 442, 465  
Swede breeding, 442  
Swede breeding, single seed descent (SSD), 433  
Sweet peas (*Lathyrus odoratus*), 36, 116, 133  
Sweet potato (*Ipomoea batatas*), 359, 371, 372, 497  
Sweet potato, breeding, 371  
Sweet potato, naturally transformed, 565  
*Synchytrium endobioticum*. *See* Potato, wart  
Syngenta, 588  
Synthetic cultivars, 291–341  
Synthetic cultivars, diploid, 332–338  
Synthetic cultivars, faba bean, 340, 341  
Synthetic cultivars, predicted performance, 332–340  
Synthetic cultivars, tetraploid, 304, 310

## T

- Tea (*Camellia sinensis*), 19  
Tetrasomic inheritance. *See* Linkage and QTL analysis  
Tetrasomic inheritance, half-sib family selection, 301  
Tetrasomic inheritance, mass selection, 296  
Tetrasomic inheritance, mixed selfing and random mating, 289

Texas Agricultural Experiment Station, 402, 474  
*Thermus aquaticus* (Taq), 99  
**TILLING** (Targeting Induced Local Lesions IN Genomes), 42, 45, 531, 535, 539, 548–552  
 Tissue culture, 256, 366, 367, 371, 372, 379, 490, 502, 543, 562, 564, 575, 584  
 Tobacco (*Nicotiana tabacum*), 19, 497, 499  
   blue mould (*Peronospora tabacina*), 499, 500  
 Tomato (*Solanum lycopersicum*), 23, 131, 154, 276, 472, 508, 623  
 Tomato, heterosis, 472  
 Tomato, hybrids, 475–476  
 Tomato, introgression, 508–509  
 Tomato, powdery mildew (*Oidium neolympicis*), 509  
 Tomato, shape, 154–155  
 Tomato, shape genes (SUN, OVATE, FASCIATED and LOCULE NUMBER), 154  
 Tomato, wild (*Solanum chilense*), 508  
 Tomato, wild (*Solanum habrochaites*), 509  
 Tomato, wild (*Solanum neorikii*), 508, 509  
 Tomato, wild (*Solanum pennellii*), 472, 508  
 Tomato, wild (*Solanum peruvianum*), 508, 509  
 Tomato, wild (*Solanum pimpinellifolium*), 508  
 Townshend, Charles, 26  
*Tradescantia paludosa*, 88  
 Transcription factor. *See* Gene expression  
 Transcriptomics. *See* Gene expression  
 Translocation breeding, 497, 518–520  
 Transposable elements (TEs). *See* Transposons  
 Transposons, 106, 107, 267, 268, 531, 553  
   Transposons, Ac, 107  
   Transposons, DNA, 107  
   Transposons, Ds, 107, 553  
   Transposons, retrotransposons, 107, 154, 267, 559  
 Transposons, *Rider* retrotransposon, 155  
 Trial designs, 168  
 Trial designs, alpha, 168  
 Trial designs, generalized lattice, 168  
 Trial designs, lattice squares, 168  
 Trial designs, precision, 166, 350, 358  
 Trial designs, randomized complete block, 167, 175, 322, 606  
 Trial designs, row and column, 168, 169  
*Trifolium pratense*. *See* Clover, red  
*Trifolium repens*. *See* Clover, white  
 Triple Test Cross (TTC). *See* Mating design  
 Triticale, 36, 41, 104, 241, 263

*Triticum aestivum*. *See* Wheat, bread  
*Triticum turgidum*. *See* Wheat, durum  
*Triticum umbellulata*, 519, 530  
*Triticum urartu*. *See* Wheat, A genome  
 Tull, Jethro, 26  
*Turnip* (*Brassica rapa*), 258

**U**  
 U, Nagaharu, 258  
*Uncinula necator*. *See* Grapevines, powdery mildew  
 United Nations (UN), 6, 49, 51, 54, 56, 57, 485  
 United Nations, Millennium Development Goals, 51, 486  
 Unreduced gametes, 92, 241–245, 247, 256  
 Unreduced gametes, first division restitution (FDR), 244–248  
 Unreduced gametes, genetic consequences, 244–248  
 Unreduced gametes, second division restitution (SDR), 244–248  
 Unreduced gametes, sexual polyploidization, 243–244  
 UPOV, 39–40  
*Ustilago scitaminea*, 265  
 UV-B radiation, 94

**V**  
 Van der Plank, J.E., 601, 603–605, 609, 613, 615, 616  
 Van Helmont, Jan Baptista, 62  
 van Leeuwenhoek, Antonie, 36  
 Vascular plants, 78  
 Vavilov, Nickolay Ivanovich, 8, 9, 15, 484  
 Vegetative reproduction, 38, 264, 274, 381, 387  
*Venturia inaequalis*. *See* Apple, scab  
 Vernalization  
   Vernalization, barley, 150–151  
   Vernalization, wheat, 150–151  
*Vicia faba*. *See* Broad bean and Faba bean  
 Vilmorin family, 33  
 Vilmorin seed company, 516  
 Virus resistance, 361, 567, 576, 582, 616  
 Virus resistance, durability, 623  
*Vitis aestivalis*, 381  
*Vitis berlandieri*, 381  
*Vitis labrusca*, 381, 382  
*Vitis riparia*, 381  
*Vitis rupestris*, 381

*Vitis vinifera*. *See* Grapevine

Vogel, Orville, 48

von Gärtnner, Karl Friedrich, 32

von Liebig, Justus, 68

von Rümker, Kurt, 33

von Sachs, Julius, 68

## W

Wageningen University DuRPh project, 583

Waldeyer, Heinrich, 37

Wallace, Henry A., 47, 48

Warwick Genetic Resources Unit  
(WGRU), 495

Water Efficient Maize for Africa (WEMA),  
52, 580

Water use efficiency (WUE), 65, 212, 221

Watson, James, 43, 44, 86

Weinberg, Wilhelm, 283

Western corn rootworm, 577

Wheat

Wheat genes ((*VRN1*, *VRN2* and *VRN3*),  
150, 151

Wheat, A genome (*Triticum urartu*), 262

Wheat, B genome (*Aegilops speltoides*),  
262, 517

Wheat, bread (*Triticum aestivum*), 262

Wheat, chromosome manipulation, 516–520

Wheat, D genome (*Aegilops tauschii*), 148, 262

Wheat, domestication, 11, 71

Wheat, durum (*Triticum turgidum*), 262

Wheat, *Fusarium* head blight resistance, 183

Wheat, homoeologous pairing, 262, 517

Wheat, (brown) leaf rust (*Puccinia triticina*),  
519, 619

Wheat, *MLO* genes, 558

Wheat, short-strawed, 48

Wheat, slow-rusting, 615, 619–621

Wheat, (black) stem rust (*Puccinia graminis*),  
604, 619

Wheat, (black) stem rust race Ug99, 619, 624

Wheat, translocation breeding, 518–520

Wheat breeding, 184, 203, 206, 440, 619

Wheat breeding, CIMMYT, 620

Wheat breeding, slow-rusting wheats, 620–621

Wheat, yellow (stripe) rust (*Puccinia  
striiformis*), 596

Wilson, J.H., 601

Wilson, James, 377

Wood, 83–84, 373

## X

*Xanthomonas axonopodis*. *See* Cassava,  
bacterial blight)

*Xanthomonas oryzae*. *See* Rice, bacterial  
blight)

## Y

Yellow monkeyflower (*Mimulus guttatus*),  
101, 102

Yield gap, 60–61, 591

Yield trials, 166–168, 170, 353, 369

Yield trials, nearest neighbour (spatial)  
analysis, 127

Younger Dryas Interval. *See* Climate

## Z

*Zea mays*. *See* Maize

*Zea mays* subspecies *parviglumis*. *See* maize,  
teosinte

*Zea mexicana*, 403