

Plant and Animal Breeding and Genetics

2nd Edition by Roos M. Zaalberg & Guillaume P. Ramstein

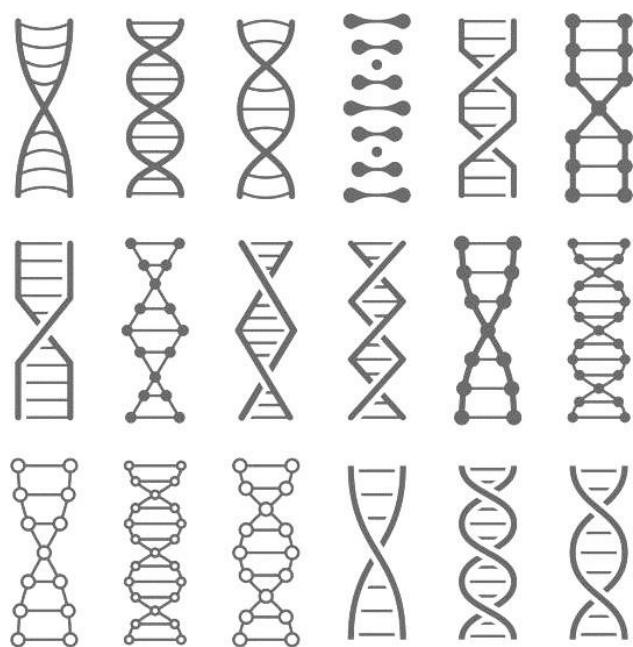


Center for Quantitative Genetics and Genomics
Aarhus Universitet
2023

TABLE OF CONTENTS

	PAGE
1. Genetics	
1.1. Lecturer's notes: Formal Genetics	
1.1.1. Roos Zaalberg 4-1-2022	6-8
1.2. Study book: <i>Klug (2019)</i>	
1.2.1. Mendelian genetics	14-22
1.2.2. Chi ² analysis	24-26
1.2.3. Patterns of heredity	28-34
1.2.4. Double heterozygous cross	36
1.2.5. Sex-determination and -chromosomes	38-51
1.2.6. Chromosomal mutations	54-63
1.2.7. Gene mutation	66-68
1.3. Study book: <i>Griffiths (2015)</i>	
1.3.1. Human pedigree analysis	75-77
1.3.2. Mapping chromosomes and recombination	81-95
1.4. Lecturer's notes: Quantitative Genetics	
1.4.1. Brief Introduction to Plant and Animal Breeding	99-101
1.4.2. Introduction to Basic Concepts in Quantitative Genetics	102-108
1.4.3. Estimation of Genetic Parameters	109-116
1.4.4. Estimation of Breeding Values	117-123
1.4.5. Estimation of Genomic Breeding Values	124-128
2. Plant breeding	
2.1. Study book: <i>Acquaah (2012)</i>	
2.1.1. Introduction to reproduction and autogamy	135-140
2.1.2. Allogamy	141-146
2.1.3. Variation: types, origin and scale	147-156
2.1.4. Plant genetic resources	157-165
2.2. Study book: <i>Brown (2014)</i>	
2.2.1. Breeding schemes	170-184
2.2.2. Selection	185-191
2.3. Scientific articles	
2.3.1. Hill (2010)	196-208
3. Livestock breeding	
3.1. Study book: <i>Simm (2020)</i>	
3.1.1. Strategies for genetic improvement	216-257
3.1.2. What affects response to selection within breeds?	260-299
3.1.3. Tools and technology in animal breeding	302-315
3.2. Study book: <i>Weber and Eskildsen (2015)</i>	
3.2.1. Pig breeding	320-340
3.3. Scientific articles	
3.3.1. Nielsen et al. (2011)	346-367

1. GENETICS



1.1 Lecturer's notes: Formal Genetics



1.1.1

Lecturer: Roos M. Zaalberg

Last edited: 4-1-2021

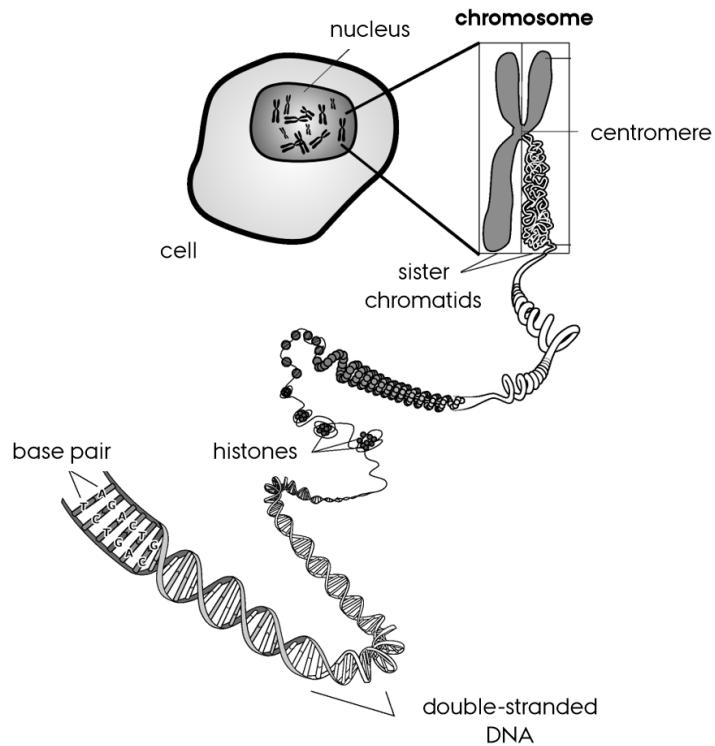


Figure 1 – Anatomy of a chromosome.

Chromosomes contain DNA, and are stored inside the cell's nucleus. They become visible after DNA duplication and condensation. During condensation, the DNA is tightly wrapped around histones. In diploid cells, chromosomes come in pairs called homologous chromosomes or homologues. In this image, we see only one homologue of a pair. The two sister chromatids of a single chromosome have the same genetic material, and are kept together with the centromere. Two homologues, on the other hand, are not identical, since one comes from father's side, and the other from mother's side.

(Image source: https://upload.wikimedia.org/wikipedia/commons/4/4c/Chromosoom_Dutch_text.png, edited by Roos Zaalberg)

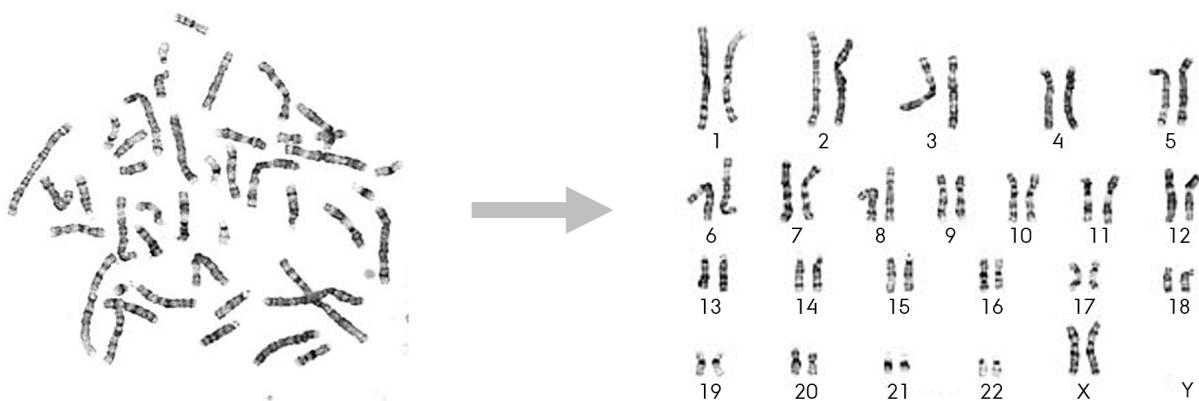


Figure 2 – Karyotype.

Karyotype of a woman, with 22 autosome pairs and one pair of X chromosomes.

(Image source: <https://www.cellgs.com/products/karyotype-service-for-live-cells.html>)

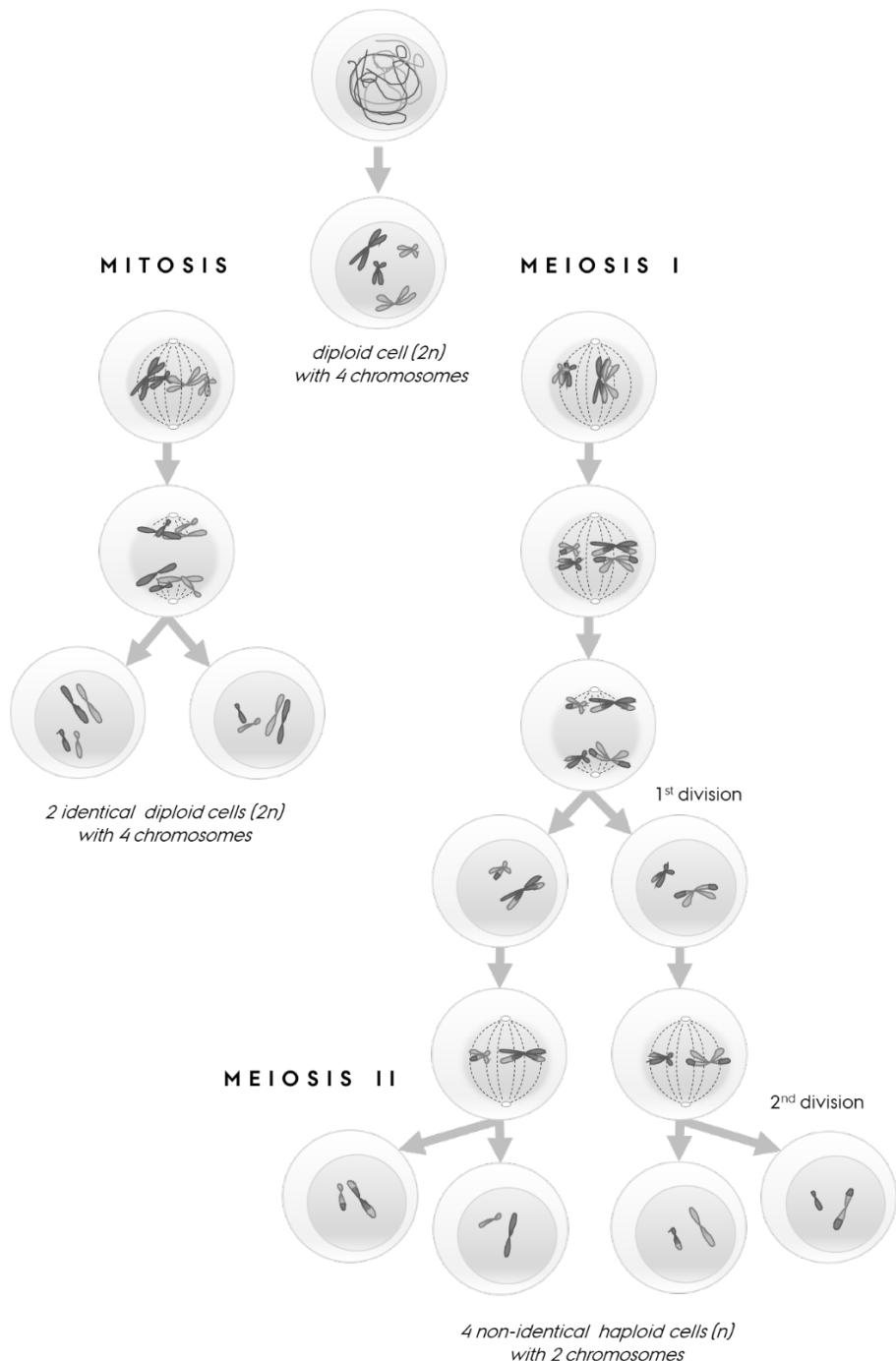


Figure 3 – Mitosis and meiosis.

Meiosis and mitosis start in the same way: stretched out DNA duplicates, and condenses. Mitosis is a cell duplication that produces two identical cells. This means that the two daughter cells will have exactly the same genetic material as the original mother cell. During meiosis, the goal is to produce a mix of different gametes (e.g. sperm and egg cells). Gametes are haploid cells, which means they have only one copy of each chromosome (n). Diploid cells, on the other hand have pairs of chromosomes ($2n$). The secret of meiosis is a process called "crossing-over", during which two homologous chromosomes pair up, and exchange genetic material. Because of crossing-over, no gamete will look the same, nor will any two humans! (Images by Roos Zaalberg)

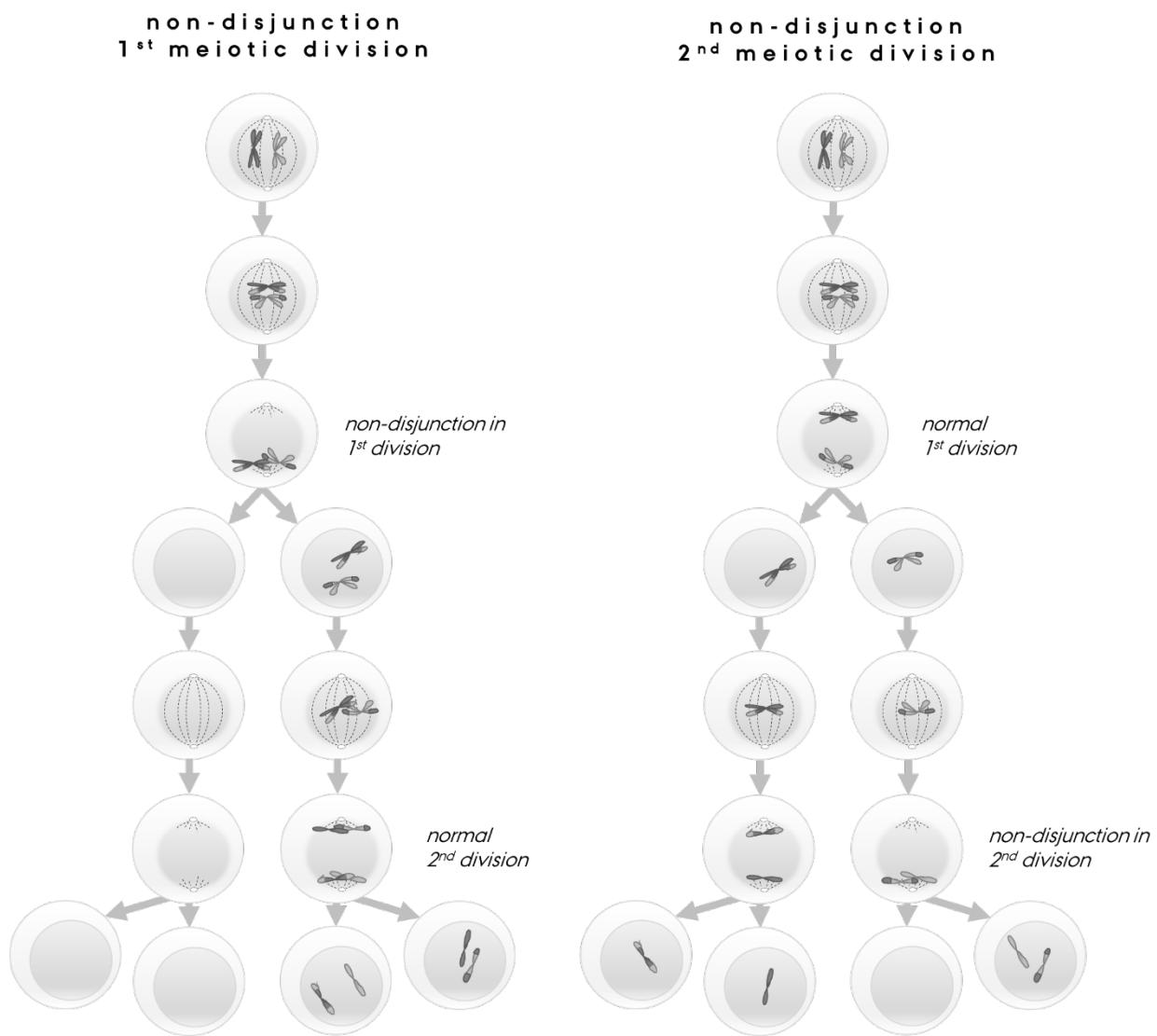


Figure 4 – Non-disjunction during meiosis.

Abnormal gametes are formed, when non-disjunction occurs during meiosis I or meiosis II. During the first meiotic division, the two homologues should each travel to their own daughter cell. However, sometimes both homologues end up in the same daughter cell. As a result, some sex cells will have no copies of this chromosome, while others will have two copies. During the second meiotic division, it can happen that the chromosome splits, but both chromatids end up in the same gamete. Gametes with two chromosomes will result in an individual with trisomy, such as Klinefelter syndrome (XXY) or Down syndrome (trisomy 21). A gamete that is missing a chromosome will result in an individual with a monomy, such as Turner syndrome (X0). Chromosomal abnormalities, such as trisomies and monomies, are often lethal. Research suggests that up to 50% of all first trimester miscarriages are due to chromosomal abnormalities.

(Images by Roos Zaalberg)

1.2 Study book



Concepts of Genetics

TWELFTH EDITION

Klug • Cummings • Spencer • Palladino • Killian

Title:	<i>Concepts of Genetics</i>
Edition:	12
Authors:	William Klug, and Michael Cummings
Year:	2019
Publisher:	Pearson

1.2.1

Klug et al. (2019)

Page 72-80

3



Mendelian Genetics

CHAPTER CONCEPTS

- Inheritance is governed by information stored in discrete unit factors called genes.
- Genes are transmitted from generation to generation on vehicles called chromosomes.
- Chromosomes, which exist in pairs in diploid organisms, provide the basis of biparental inheritance.
- During gamete formation, chromosomes are distributed according to postulates first described by Gregor Mendel, based on his nineteenth-century research with the garden pea.
- Mendelian postulates prescribe that homologous chromosomes segregate from one another and assort independently with other segregating homologs during gamete formation.
- Genetic ratios, expressed as probabilities, are subject to chance deviation and may be evaluated statistically.
- The analysis of pedigrees allows predictions concerning the genetic nature of human traits.

Although inheritance of biological traits has been recognized for thousands of years, the first significant insights into how it takes place only occurred about 150 years ago. In 1866, Gregor Johann Mendel published the results of a series of experiments that would lay the

Gregor Johann Mendel, who in 1866 put forward the major postulates of transmission genetics as a result of experiments with the garden pea.

foundation for the formal discipline of genetics. Mendel's work went largely unnoticed until the turn of the twentieth century, but eventually, the concept of the gene as a distinct hereditary unit was established. Since then, the ways in which genes, as segments of chromosomes, are transmitted to offspring and control traits have been clarified. Research continued unabated throughout the twentieth century and into the present—indeed, studies in genetics, most recently at the molecular level, have remained at the forefront of biological research since the early 1900s.

When Mendel began his studies of inheritance using *Pisum sativum*, the garden pea, chromosomes and the role and mechanism of meiosis were totally unknown. Nevertheless, he determined that discrete units of inheritance exist and predicted their behavior in the formation of gametes. Subsequent investigators, with access to cytological data, were able to relate their own observations of chromosome behavior during meiosis and Mendel's principles of inheritance. Once this correlation was recognized, Mendel's postulates were accepted as the basis for the study of what is known as **transmission genetics**—how genes are transmitted from parents to offspring. These principles were derived directly from Mendel's experimentation.

3.1 Mendel Used a Model Experimental Approach to Study Patterns of Inheritance

Johann Mendel was born in 1822 to a peasant family in the Central European village of Heinzendorf. An excellent student in high school, he studied philosophy for several years afterward and in 1843, taking the name Gregor, was admitted to the Augustinian Monastery of St. Thomas in Brno, now part of the Czech Republic. In 1849, he was relieved of pastoral duties, and from 1851 to 1853, he attended the University of Vienna, where he studied physics and botany. He returned to Brno in 1854, where he taught physics and natural science for the next 16 years. Mendel received support from the monastery for his studies and research throughout his life.

In 1856, Mendel performed his first set of hybridization experiments with the garden pea, launching the research phase of his career. His experiments continued until 1868, when he was elected abbot of the monastery. Although he retained his interest in genetics, his new responsibilities demanded most of his time. In 1884, Mendel died of a

kidney disorder. The local newspaper paid him the following tribute:

His death deprives the poor of a benefactor, and mankind at large of a man of the noblest character, one who was a warm friend, a promoter of the natural sciences, and an exemplary priest.

Mendel first reported the results of some simple genetic crosses between certain strains of the garden pea in 1865. Although his was not the first attempt to provide experimental evidence pertaining to inheritance, Mendel's success where others had failed can be attributed, at least in part, to his elegant experimental design and analysis.

Mendel showed remarkable insight into the methodology necessary for good experimental biology. First, he chose an organism that was easy to grow and to hybridize artificially. The pea plant is self-fertilizing in nature, but it is easy to cross-breed experimentally. It reproduces well and grows to maturity in a single season. Mendel followed seven visible features (we refer to them as characters, or characteristics), each represented by two contrasting forms, or **traits** (Figure 3.1). For the character stem height, for example, he experimented with the traits *tall* and *dwarf*. He selected

Character	Contrasting traits		F ₁ results	F ₂ results	F ₂ ratio
Seed shape	round/wrinkled		all round	5474 round 1850 wrinkled	2.96:1
Seed color	yellow/green		all yellow	6022 yellow 2001 green	3.01:1
Pod shape	full/constricted		all full	882 full 299 constricted	2.95:1
Pod color	green/yellow		all green	428 green 152 yellow	2.82:1
Flower color	violet/white		all violet	705 violet 224 white	3.15:1
Flower position	axial/terminal		all axial	651 axial 207 terminal	3.14:1
Stem height	tall/dwarf		all tall	787 tall 277 dwarf	2.84:1

FIGURE 3.1 Seven pairs of contrasting traits and the results of Mendel's seven monohybrid crosses of the garden pea (*Pisum sativum*). In each case, pollen derived from plants exhibiting one trait was used to fertilize the ova of plants

exhibiting the other trait. In the F₁ generation, one of the two traits was exhibited by all plants. The contrasting trait reappeared in approximately 1/4 of the F₂ plants.

six other contrasting pairs of traits involving seed shape and color, pod shape and color, and flower color and position. From local seed merchants, Mendel obtained true-breeding strains, those in which each trait appeared unchanged generation after generation in self-fertilizing plants.

There were several other reasons for Mendel's success. In addition to his choice of a suitable organism, he restricted his examination to one or very few pairs of contrasting traits in each experiment. He also kept accurate quantitative records, a necessity in genetic experiments. From the analysis of his data, Mendel derived certain postulates that have become the principles of transmission genetics.

The results of Mendel's experiments went unappreciated until the turn of the century, well after his death. However, once Mendel's publications were rediscovered by geneticists investigating the function and behavior of chromosomes, the implications of his postulates were immediately apparent. He had discovered the basis for the transmission of hereditary traits!

3.2 The Monohybrid Cross Reveals How One Trait Is Transmitted from Generation to Generation

Mendel's simplest crosses involved only one pair of contrasting traits. Each such experiment is called a **monohybrid cross**. A monohybrid cross is made by mating true-breeding individuals from two parent strains, each exhibiting one of the two contrasting forms of the character under study. Initially, we examine the first generation of offspring of such a cross, and then we consider the offspring of **selfing**, that is, of self-fertilization of individuals from this first generation. The original parents constitute the **P₁**, or **parental generation**; their offspring are the **F₁**, or **first filial generation**; the individuals resulting from the selfed F₁ generation are the **F₂**, or **second filial generation**; and so on.

The cross between true-breeding pea plants with tall stems and dwarf stems is representative of Mendel's monohybrid crosses. *Tall* and *dwarf* are contrasting traits of the character of stem height. Unless tall or dwarf plants are crossed together or with another strain, they will undergo self-fertilization and breed true, producing their respective traits generation after generation. However, when Mendel crossed tall plants with dwarf plants, the resulting F₁ generation consisted of only tall plants. When members of the F₁ generation were selfed, Mendel observed that 787 of 1064 F₂ plants were tall, while 277 of 1064 were dwarf. Note that in this cross (Figure 3.1), the dwarf trait disappeared in the F₁ generation, only to reappear in the F₂ generation.

Genetic data are usually expressed and analyzed as ratios. In this particular example, many identical P₁ crosses

were made and many F₁ plants—all tall—were produced. As noted, of the 1064 F₂ offspring, 787 were tall and 277 were dwarf—a ratio of approximately 2.8:1.0, or about 3:1.

Mendel made similar crosses between pea plants exhibiting each of the other pairs of contrasting traits; the results of these crosses are shown in Figure 3.1. In every case, the outcome was similar to the tall/dwarf cross just described. For the character of interest, all F₁ offspring expressed the same trait exhibited by one of the parents, but in the F₂ offspring, an approximate ratio of 3:1 was obtained. That is, three-fourths looked like the F₁ plants, while one-fourth exhibited the contrasting trait, which had disappeared in the F₁ generation.

We note one further aspect of Mendel's monohybrid crosses. In each cross, the F₁ and F₂ patterns of inheritance were similar regardless of which P₁ plant served as the source of pollen (sperm) and which served as the source of the ovum (egg). The crosses could be made either way—pollination of dwarf plants by tall plants, or vice versa. Crosses made in both these ways are called **reciprocal crosses**. Therefore, the results of Mendel's monohybrid crosses were not sex dependent.

To explain these results, Mendel proposed the existence of particulate *unit factors* for each trait. He suggested that these factors serve as the basic units of heredity and are passed unchanged from generation to generation, determining various traits expressed by each individual plant. Using these general ideas, Mendel proceeded to hypothesize precisely how such factors could account for the results of the monohybrid crosses.

Mendel's First Three Postulates

Using the consistent pattern of results in the monohybrid crosses, Mendel derived the following three postulates, or principles, of inheritance.

1. UNIT FACTORS IN PAIRS

Genetic characters are controlled by unit factors existing in pairs in individual organisms.

In the monohybrid cross involving tall and dwarf stems, a specific **unit factor** exists for each trait. Each diploid individual receives one factor from each parent. Because the factors occur in pairs, three combinations are possible: two factors for tall stems, two factors for dwarf stems, or one of each factor. Every individual possesses one of these three combinations, which determines stem height.

2. DOMINANCE/RECESSIVENESS

When two unlike unit factors responsible for a single character are present in a single individual, one unit factor is dominant to the other, which is said to be recessive.

In each monohybrid cross, the trait expressed in the F₁ generation is controlled by the dominant unit factor. The trait not expressed is controlled by the recessive unit factor. The terms dominant and recessive are also

used to designate traits. In this case, tall stems are said to be dominant over recessive dwarf stems.

3. SEGREGATION

During the formation of gametes, the paired unit factors separate, or segregate, randomly so that each gamete receives one or the other with equal likelihood.

If an individual contains a pair of like unit factors (e.g., both specific for tall), then all its gametes receive one of that same kind of unit factor (in this case, tall). If an individual contains unlike unit factors (e.g., one for tall and one for dwarf), then each gamete has a 50 percent probability of receiving either the tall or the dwarf unit factor.

These postulates provide a suitable explanation for the results of the monohybrid crosses. Let's use the tall/dwarf cross to illustrate. Mendel reasoned that P_1 tall plants contained identical paired unit factors, as did the P_1 dwarf plants. The gametes of tall plants all receive one tall unit factor as a result of segregation. Similarly, the gametes of dwarf plants all receive one dwarf unit factor. Following fertilization, all F_1 plants receive one unit factor from each parent—a tall factor from one and a dwarf factor from the other—reestablishing the paired relationship, but because tall is dominant to dwarf, all F_1 plants are tall.

When F_1 plants form gametes, the postulate of segregation demands that each gamete randomly receives either the tall or dwarf unit factor. Following random fertilization events during F_1 selfing, four F_2 combinations will result with equal frequency:

1. tall/tall
2. tall/dwarf
3. dwarf/tall
4. dwarf/dwarf

Combinations (1) and (4) will clearly result in tall and dwarf plants, respectively. According to the postulate of dominance/recessiveness, combinations (2) and (3) will both yield tall plants. Therefore, the F_2 is predicted to consist of 3/4 tall and 1/4 dwarf, or a ratio of 3:1. This is approximately what Mendel observed in his cross between tall and dwarf plants. A similar pattern was observed in each of the other monohybrid crosses (Figure 3.1).

Modern Genetic Terminology

To analyze the monohybrid cross and Mendel's first three postulates, we must first introduce several new terms as well as a symbol convention for the unit factors. Traits such as tall or dwarf are physical expressions of the information contained in unit factors. The physical expression of a trait is the **phenotype** of the individual. Mendel's unit factors represent

units of inheritance called **genes** by modern geneticists. For any given character, such as plant height, the phenotype is determined by alternative forms of a single gene, called **alleles**. For example, the unit factors representing tall and dwarf are alleles determining the height of the pea plant.

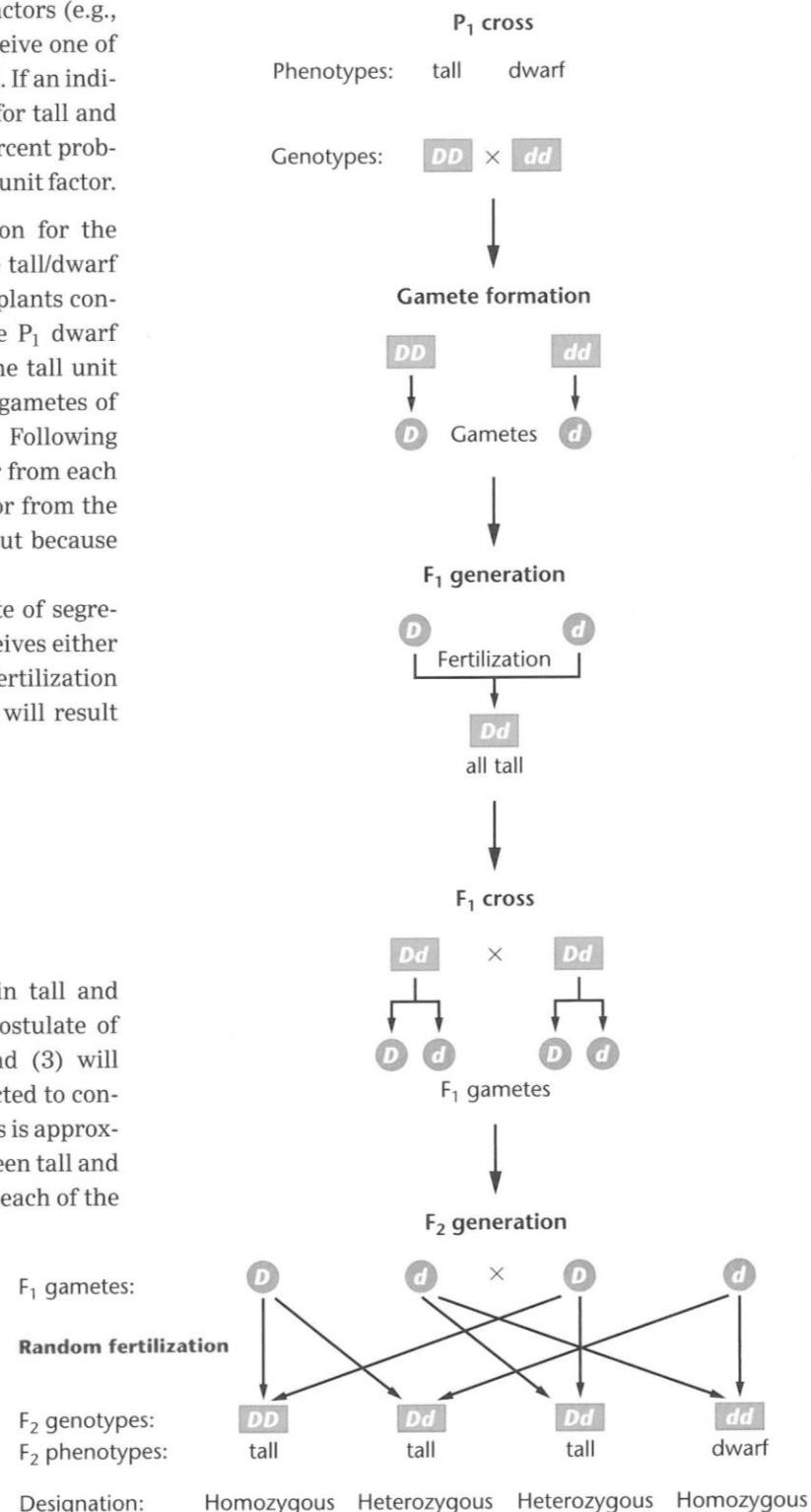


FIGURE 3.2 The monohybrid cross between tall (D) and dwarf (d) pea plants. Individuals are shown in rectangles, and gametes are shown in circles.

Geneticists have several different systems for using symbols to represent genes. Later in the text (see Chapter 4), we will review a number of these conventions, but for now, we will adopt one to use consistently throughout this chapter. According to this convention, the first letter of the recessive trait symbolizes the character in question; in lowercase italic, it designates the allele for the recessive trait, and in uppercase italic, it designates the allele for the dominant trait. Thus for Mendel's pea plants, we use *d* for the dwarf allele and *D* for the tall allele. When alleles are written in pairs to represent the two unit factors present in any individual (*DD*, *Dd*, or *dd*), the resulting symbol is called the **genotype**. The genotype designates the genetic makeup of an individual for the trait or traits it describes, whether the individual is haploid or diploid. By reading the genotype, we know the phenotype of the individual: *DD* and *Dd* are tall, and *dd* is dwarf. When both alleles are the same (*DD* or *dd*), the individual is **homozygous** for the trait, or a **homozygote**; when the alleles are different (*Dd*), we use the terms **heterozygous** and **heterozygote**. These symbols and terms are used in Figure 3.2 to describe the monohybrid cross, as discussed on page 75.

Punnett Squares

The genotypes and phenotypes resulting from combining gametes during fertilization can be easily visualized by constructing a diagram called a **Punnett square**, named after the person who first devised this approach, Reginald C. Punnett. Figure 3.3 illustrates this method of analysis for our $F_1 \times F_1$ monohybrid cross. Each of the possible gametes is assigned a column or a row; the vertical columns represent those of the female parent, and the horizontal rows represent those of the male parent. After assigning the gametes to the rows and columns, we predict the new generation by entering the male and female gametic information into each box and thus producing every possible resulting genotype. By filling out the Punnett square, we are listing all possible random fertilization events. The genotypes and phenotypes of all potential offspring are ascertained by reading the combinations in the boxes.

The Punnett square method is particularly useful when you are first learning about genetics and how to solve genetics problems. Note the ease with which the 3:1 phenotypic ratio and the 1:2:1 genotypic ratio may be derived for the F_2 generation in Figure 3.3.

The Testcross: One Character

Tall plants produced in the F_2 generation are predicted to have either the *DD* or the *Dd* genotype. You might ask if there is a way to distinguish the genotype. Mendel devised a rather simple method that is still used today to discover the genotype of plants and animals: the **testcross**. The organism expressing the dominant phenotype but having an unknown genotype is crossed with a known

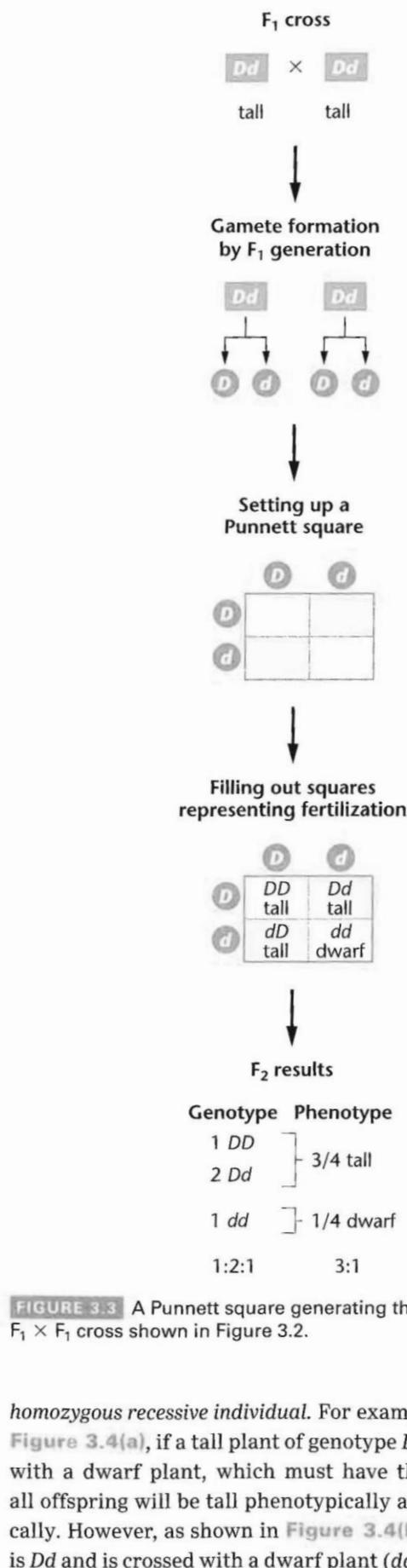


FIGURE 3.3 A Punnett square generating the F_2 ratio of the $F_1 \times F_1$ cross shown in Figure 3.2.

homozygous recessive individual. For example, as shown in Figure 3.4(a), if a tall plant of genotype *DD* is testcrossed with a dwarf plant, which must have the *dd* genotype, all offspring will be tall phenotypically and *Dd* genotypically. However, as shown in Figure 3.4(b), if a tall plant is *Dd* and is crossed with a dwarf plant (*dd*), then one-half

NOW SOLVE THIS

3.1 Pigeons may exhibit a checkered or plain color pattern. In a series of controlled matings, the following data were obtained.

P ₁ Cross	F ₁ Progeny	
	Checkered	Plain
(a) checkered × checkered	36	0
(b) checkered × plain	38	0
(c) plain × plain	0	35

Then F₁ offspring were selectively mated with the following results. (The P₁ cross giving rise to each F₁ pigeon is indicated in parentheses.)

F ₁ × F ₁ Crosses	F ₂ Progeny	
	Checkered	Plain
(d) checkered (a) × plain (c)	34	0
(e) checkered (b) × plain (c)	17	14
(f) checkered (b) × checkered (b)	28	9
(g) checkered (a) × checkered (b)	39	0

How are the checkered and plain patterns inherited? Select and assign symbols for the genes involved, and determine the genotypes of the parents and offspring in each cross.

HINT: This problem asks you to analyze the data produced from several crosses involving pigeons and to determine the mode of inheritance and the genotypes of the parents and offspring in a number of instances. The key to its solution is to first determine whether or not this is a monohybrid cross. To do so, convert the data to ratios that are characteristic of Mendelian crosses. In the case of this problem, ask first whether any of the F₂ ratios match Mendel's 3:1 monohybrid ratio. If so, the second step is to determine which trait is dominant and which is recessive.

of the offspring will be tall (*Dd*) and the other half will be dwarf (*dd*). Therefore, a 1:1 tall/dwarf ratio demonstrates the heterozygous nature of the tall plant of unknown genotype. The results of the testcross reinforced Mendel's conclusion that separate unit factors control traits.

3.3 Mendel's Dihybrid Cross Generated a Unique F₂ Ratio

As a natural extension of the monohybrid cross, Mendel also designed experiments in which he examined two characters simultaneously. Such a cross, involving two pairs of contrasting traits, is a **dihybrid cross**, or a **two-factor cross**. For example, if pea plants having yellow seeds that are round were bred with those having green seeds that are wrinkled, the results shown in Figure 3.5 would occur: the

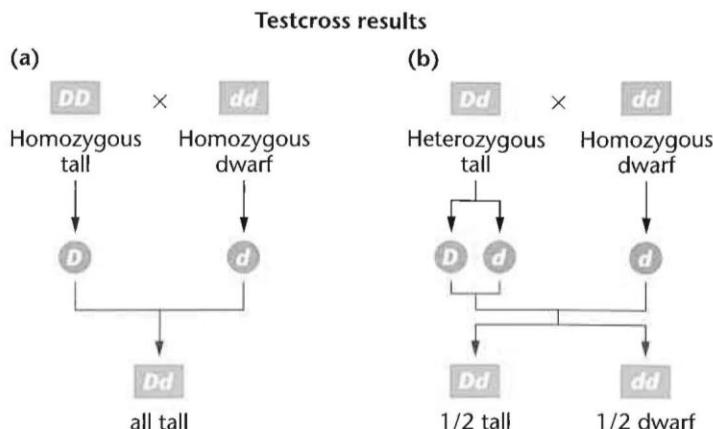


FIGURE 3.4 Testcross of a single character. In (a), the tall parent is homozygous, but in (b), the tall parent is heterozygous. The genotype of each tall P₁ plant can be determined by examining the offspring when each is crossed with the homozygous recessive dwarf plant.

F₁ offspring would all be yellow and round. It is therefore apparent that yellow is dominant to green and that round is dominant to wrinkled. When the F₁ individuals are selfed, approximately 9/16 of the F₂ plants express the yellow and round traits, 3/16 express yellow and wrinkled, 3/16 express green and round, and 1/16 express green and wrinkled.

A variation of this cross is also shown in Figure 3.5. Instead of crossing one P₁ parent with both dominant traits (yellow, round) to one with both recessive traits (green, wrinkled), plants with yellow, wrinkled seeds are crossed with those with green, round seeds. In spite of the change in the P₁ phenotypes, both the F₁ and F₂ results remain unchanged. Why this is so will become clear below.

Mendel's Fourth Postulate: Independent Assortment

We can most easily understand the results of a dihybrid cross if we consider it theoretically as consisting of two monohybrid crosses conducted separately. Think of the two sets of traits as being inherited independently of each other; that is, the chance of any plant having yellow or green seeds is not at all influenced by the chance that this plant will have round or wrinkled seeds. Thus, because yellow is dominant to green, all F₁ plants in the first theoretical cross would have yellow seeds. In the second theoretical cross, all F₁ plants would have round seeds because round is dominant to wrinkled. When Mendel examined the F₁ plants of the dihybrid cross, all were yellow and round, as our theoretical crosses predict.

The predicted F₂ results of the first cross are 3/4 yellow and 1/4 green. Similarly, the second cross would yield 3/4 round and 1/4 wrinkled. Figure 3.5 shows that in the dihybrid cross, 12/16 F₂ plants are yellow, while 4/16 are green, exhibiting the expected 3:1 (3/4:1/4) ratio. Similarly, 12/16 of all F₂ plants have round seeds, while 4/16 have wrinkled seeds, again revealing the 3:1 ratio.

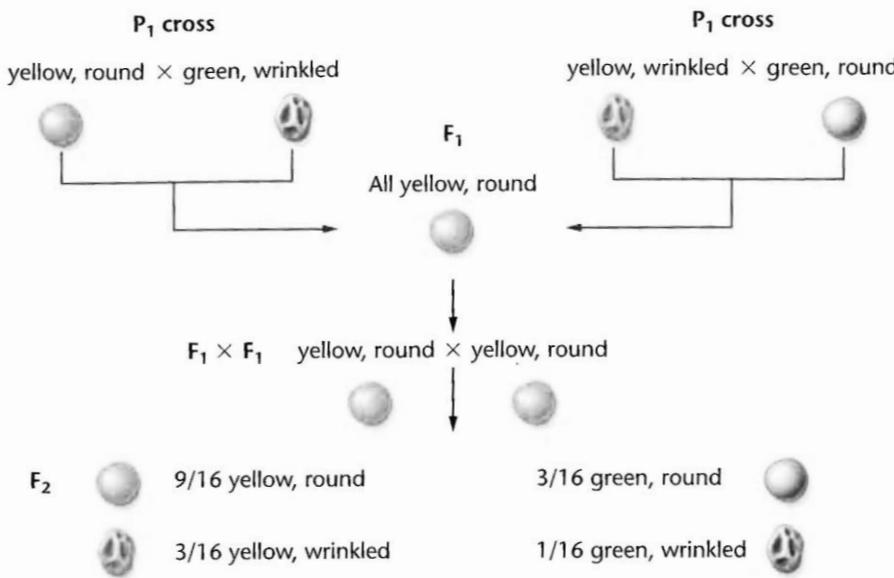


FIGURE 3.5 F₁ and F₂ results of Mendel's dihybrid crosses in which the plants on the top left with yellow, round seeds are crossed with plants having green, wrinkled seeds, and the plants on the top right with yellow, wrinkled seeds are crossed with plants having green, round seeds.

These numbers demonstrate that the two pairs of contrasting traits are inherited independently, so we can predict the frequencies of all possible F₂ phenotypes by applying the **product law** of probabilities: *the probability of two or more independent events occurring simultaneously is equal to the product of their individual probabilities*. For example, the probability of an F₂ plant having yellow and round seeds is (3/4)(3/4), or 9/16, because 3/4 of all F₂ plants should be yellow and 3/4 of all F₂ plants should be round.

In a like manner, the probabilities of the other three F₂ phenotypes can be calculated: yellow (3/4) and wrinkled (1/4) are predicted to be present together 3/16 of the time; green (1/4) and round (3/4) are predicted 3/16 of the time; and green (1/4) and wrinkled (1/4) are predicted 1/16 of the time. These calculations are shown in Figure 3.6.

It is now apparent why the F₁ and F₂ results are identical whether the initial cross is yellow, round plants bred with green, wrinkled plants, or whether yellow, wrinkled plants are bred with green, round plants. In both crosses, the F₁ genotype of all offspring is identical. As a result, the F₂ generation is also identical in both crosses.

On the basis of similar results in numerous dihybrid crosses, Mendel proposed a fourth postulate:

4. INDEPENDENT ASSORTMENT

During gamete formation, segregating pairs of unit factors assort independently of each other.

This postulate stipulates that segregation of any pair of unit factors occurs independently of all others. As a result of random segregation, each gamete receives one member of every pair of unit factors. For one pair, whichever unit factor is received does not influence the outcome of segregation of any other pair. Thus, according to the postulate of independent assortment, all possible combinations of gametes should be formed in equal frequency.

The Punnett square in Figure 3.7 shows how independent assortment works in the formation of the F₂ generation. Examine the formation of gametes by the F₁ plants; segregation prescribes that every gamete receives either a G or g allele and a W or w allele. Independent assortment stipulates that all four combinations (GW, Gw, gW, and gw) will be formed with equal probabilities.

F ₁ yellow, round × yellow, round		
F ₂	Of all offspring	Combined probabilities
3/4 are yellow	↓ 3/4 are round and 1/4 are wrinkled	→ (3/4)(3/4) = 9/16 yellow, round → (3/4)(1/4) = 3/16 yellow, wrinkled
1/4 are green	↓ 3/4 are round and 1/4 are wrinkled	→ (1/4)(3/4) = 3/16 green, round → (1/4)(1/4) = 1/16 green, wrinkled

FIGURE 3.6 Computation of the combined probabilities of each F₂ phenotype for two independently inherited characters. The probability of each plant being yellow or green is independent of the probability of it bearing round or wrinkled seeds.

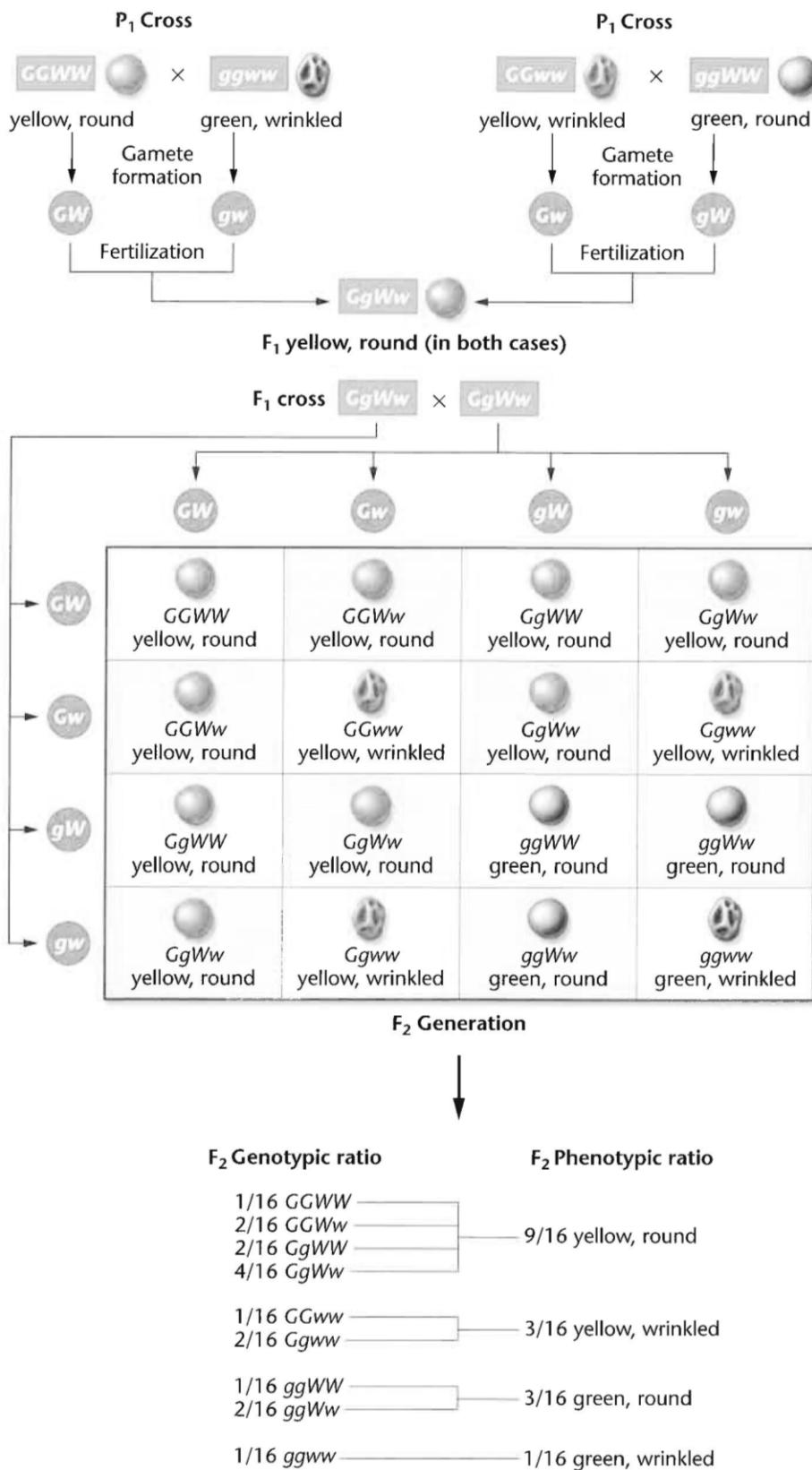


FIGURE 3.7 Analysis of the dihybrid crosses shown in Figure 3.5. The F₁ heterozygous plants are self-fertilized to produce an F₂ generation, which is computed using a Punnett square. Both the phenotypic and genotypic F₂ ratios are shown.

In every $F_1 \times F_1$ fertilization event, each zygote has an equal probability of receiving one of the four combinations from each parent. If many offspring are produced, 9/16 have yellow, round seeds, 3/16 have yellow, wrinkled seeds, 3/16 have green, round seeds, and 1/16 have green, wrinkled seeds, yielding what is designated as **Mendel's 9:3:3:1 dihybrid ratio**. This is an ideal ratio based on probability events involving segregation, independent assortment, and random fertilization. Because of deviation due strictly to chance,

particularly if small numbers of offspring are produced, actual results are highly unlikely to match the ideal ratio.

1.2.2

Klug et al. (2019)

Page 85-87

3.8 Chi-Square Analysis Evaluates the Influence of Chance on Genetic Data

Mendel's 3:1 monohybrid and 9:3:3:1 dihybrid ratios are hypothetical predictions based on the following assumptions: (1) each allele is dominant or recessive, (2) segregation is unimpeded, (3) independent assortment occurs, and (4) fertilization is random. The final two assumptions are influenced by chance events and therefore are subject to random fluctuation. This concept of **chance deviation** is most easily illustrated by tossing a single coin numerous times and recording the number of heads and tails observed. In each toss, there is a probability of 1/2 that a head will occur and a probability of 1/2 that a tail will occur. Therefore, the expected ratio of many tosses is 1/2:1/2, or 1:1. If a coin is tossed 1000 times, usually *about* 500 heads and 500 tails will be observed. Any reasonable fluctuation from this hypothetical ratio (e.g., 486 heads and 514 tails) is attributed to chance.

As the total number of tosses is reduced, the impact of chance deviation increases. For example, if a coin is tossed only four times, you would not be too surprised if all four tosses resulted in only heads or only tails. For 1000 tosses, however, 1000 heads or 1000 tails would be most unexpected. In fact, you might believe that such a result would be impossible. Actually, all heads or all tails in 1000 tosses can be predicted to occur with a probability of $(1/2)^{1000}$. Since $(1/2)^{20}$ is less than one in a million times, an event occurring with a probability as small as $(1/2)^{1000}$ is virtually impossible. Two major points to keep in mind when predicting or analyzing genetic outcomes are:

1. The outcomes of independent assortment and fertilization, like coin tossing, are subject to random fluctuations from their predicted occurrences as a result of chance deviation.
2. As the sample size increases, the average deviation from the expected results decreases. Therefore, a larger sample size diminishes the impact of chance deviation on the final outcome.

Chi-Square Calculations and the Null Hypothesis

In genetics, being able to evaluate observed deviation is a crucial skill. When we assume that data will fit a given ratio such as 1:1, 3:1, or 9:3:3:1, we establish what is called the **null hypothesis (H_0)**. It is so named because the hypothesis assumes that there is *no real difference* between the *measured values* (or ratio) and the *predicted values* (or ratio). Any apparent difference can be attributed purely to chance. The validity of the null hypothesis for a given set of data is measured using statistical analysis. Depending on the results of this analysis, the null hypothesis may either (1) *be rejected* or (2) *fail to be rejected*. If it is rejected, the observed deviation from the expected result is judged not to be attributable to chance alone. In this case, the null hypothesis and the underlying assumptions leading to it must be reexamined. If the null hypothesis fails to be rejected, any observed deviations are attributed to chance.

One of the simplest statistical tests for assessing the goodness of fit of the null hypothesis is **chi-square (χ^2) analysis**. This test takes into account the observed deviation in each component of a ratio (from what was expected) as well as the sample size and reduces them to a single numerical value. The value for χ^2 is then used to estimate how frequently the observed deviation can be expected to occur strictly as a result of chance. The formula used in chi-square analysis is

$$\chi^2 = \Sigma \frac{(o - e)^2}{e}$$

where o is the observed value for a given category, e is the expected value for that category, and Σ (the Greek letter sigma) represents the sum of the calculated values for each category in the ratio. Because $(o - e)$ is the deviation (d) in each case, the equation reduces to

$$\chi^2 = \Sigma \frac{d^2}{e}$$

Table 3.1(a) shows the steps in the χ^2 calculation for the F_2 results of a hypothetical monohybrid cross. To analyze the data obtained from this cross, work from left to right across the table, verifying the calculations as appropriate. Note that regardless of whether the deviation d is positive or negative, d^2 always becomes positive after the number is squared. In **Table 3.1(b)** F_2 results of a hypothetical dihybrid cross are analyzed. Make sure that you understand how each number was calculated in this example.

The final step in chi-square analysis is to interpret the χ^2 value. To do so, you must initially determine a value called the **degrees of freedom (df)**, which is equal to $n - 1$, where n is the number of different categories into which the data are divided, in other words, the number of possible outcomes. For the 3:1 ratio, $n = 2$, so $df = 1$. For the 9:3:3:1 ratio, $n = 4$ and $df = 3$. Degrees of freedom must be taken into account because the greater the number of categories, the more deviation is expected as a result of chance.

Once you have determined the degrees of freedom, you can interpret the χ^2 value in terms of a corresponding **probability value (p)**. Since this calculation is complex, we usually take the p value from a standard table or graph. **Figure 3.11** shows a wide range of χ^2 values and the corresponding p values for various degrees of freedom in both a graph and a table. Let's use the graph to explain how to determine the p value. The caption for **Figure 3.11(b)** explains how to use the table.

To determine p using the graph, execute the following steps:

- Locate the χ^2 value on the abscissa (the horizontal axis, or x -axis).
- Draw a vertical line from this point up to the line on the graph representing the appropriate df .

TABLE 3.1 Chi-Square Analysis

(a) Monohybrid

Cross Expected Ratio	Observed (o)	Expected (e)	Deviation ($o - e = d$)	Deviation 2	d^2/e
3/4	740	3/4(1000) = 750	740 - 750 = -10	(-10) 2 = 100	100/750 = 0.13
1/4	260	1/4(1000) = 250	260 - 250 = +10	(+10) 2 = 100	100/250 = 0.40
	Total = 1000				$\chi^2 = 0.53$ $p = 0.48$

(b) Dihybrid

Cross Expected Ratio	Observed (o)	Expected (e)	Deviation ($o - e = d$)	Deviation 2	d^2/e
9/16	587	567	+20	400	0.71
3/16	197	189	+8	64	0.34
3/16	168	189	-21	441	2.33
1/16	56	63	-7	49	0.78
	Total = 1008				$\chi^2 = 4.16$ $p = 0.26$

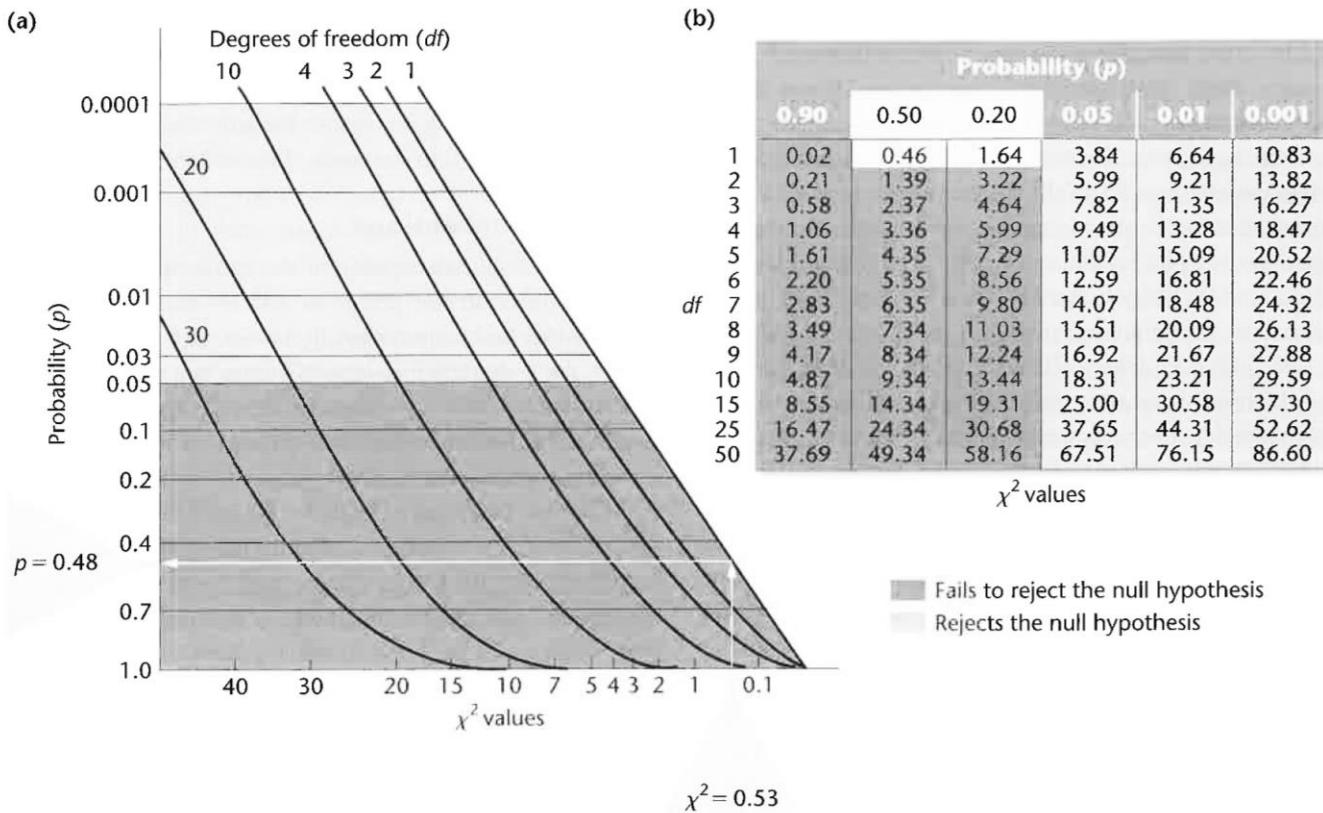


FIGURE 3.11 (a) Graph for converting χ^2 values to p values. (b) Table of χ^2 values for selected values of df and p . χ^2 values that lead to a p value of 0.05 or greater (darker blue areas) justify failure to reject the null hypothesis. Values leading to a p value of less than 0.05 (lighter blue areas) justify

rejecting the null hypothesis. For example, the table in part (b) shows that for $\chi^2 = 0.53$ with 1 degree of freedom, the corresponding p value is between 0.20 and 0.50. The graph in (a) gives a more precise p value of 0.48 by interpolation. Thus, we fail to reject the null hypothesis.

3. From there, extend a horizontal line to the left until it intersects the ordinate (the vertical axis, or y -axis).
4. Estimate, by interpolation, the corresponding p value.

We used these steps for the monohybrid cross in Table 3.1(a) to estimate the p value of 0.48, as shown in Figure 3.11(a). Now try this method to see if you can determine the p value for the dihybrid cross [Table 3.1(b)]. Since the χ^2 value is 4.16 and $df = 3$, an approximate p value is 0.26. Checking this result in the table confirms that p values for both the monohybrid and dihybrid crosses are between 0.20 and 0.50.

1.2.3

Klug et al. (2019)

Page 100-106

4.2 Geneticists Use a Variety of Symbols for Alleles

Earlier in the text, we learned a standard convention used to symbolize alleles for very simple Mendelian traits (see Chapter 3). The initial letter of the name of a recessive trait, lowercased and italicized, denotes the recessive allele, and the same letter in uppercase refers to the dominant allele. Thus, in the case of tall and dwarf, where dwarf is recessive, *D* and *d* represent the alleles responsible for these respective traits. Mendel used upper- and lowercase letters such as these to symbolize his unit factors.

Another useful system was developed in genetic studies of the fruit fly *Drosophila melanogaster* to discriminate between wild-type and mutant traits. This system uses the initial letter, or a combination of several letters, from the name of the mutant trait. If the trait is recessive, lowercase is used; if it is dominant, uppercase is used. The contrasting wild-type trait is denoted by the same letters, but with a superscript +. For example, *ebony* is a recessive body color mutation in *Drosophila*. The normal wild-type body color is gray. Using this system, we denote *ebony* by the symbol *e*, while gray is denoted by *e*⁺. The responsible locus may be occupied by either the wild-type allele (*e*⁺) or the mutant allele (*e*). A diploid fly may thus exhibit one of three possible genotypes (the two phenotypes are indicated parenthetically):

<i>e</i> ⁺ / <i>e</i> ⁺	gray homozygote (wild type)
<i>e</i> ⁺ / <i>e</i>	gray heterozygote (wild type)
<i>e</i> / <i>e</i>	ebony homozygote (mutant)

The slash between the letters indicates that the two allele designations represent the same locus on two homologous chromosomes. If we instead consider a mutant allele that is dominant to the normal wild-type allele, such as *Wrinkled wing* in *Drosophila*, the three possible genotypes are *Wr/Wr*, *Wr/Wr*⁺, and *Wr*⁺/*Wr*⁺. The initial two genotypes express the mutant wrinkled-wing phenotype.

One advantage of this system is that further abbreviation can be used when convenient: The wild-type allele may simply be denoted by the + symbol. With *ebony* as an example, the designations of the three possible genotypes become the following:

<i>+/+</i>	gray homozygote (wild type)
<i>/e</i>	gray heterozygote (wild type)
<i>e/e</i>	ebony homozygote (mutant)

Another variation is utilized when no dominance exists between alleles (a situation we will explore in Section 4.3). We simply use uppercase letters and superscripts to denote alternative alleles (e.g., *R*¹ and *R*², *L*^M and *L*^N, and *I*^A and *I*^B).

Many diverse systems of genetic nomenclature are used to identify genes in various organisms. Usually, the symbol selected reflects the function of the gene or even a disorder caused by a mutant gene. For example, in yeast, *cdk* is the abbreviation for the cyclin-dependent kinase gene, whose product is involved in the cell-cycle regulation mechanism (discussed in Chapter 2). In bacteria, *leu*⁻ refers to a mutation that interrupts the biosynthesis of the amino acid leucine, and the wild-type gene is designated *leu*⁺. The symbol *dnaA* represents a bacterial gene involved in DNA replication (and *DnaA*, without italics, is the protein made by that gene). In humans, italicized capital letters are used to name genes: *BRCA1* represents one of the genes associated with susceptibility to breast cancer. Although these different systems may seem complex, they are useful ways to symbolize genes.

4.3 Neither Allele Is Dominant in Incomplete, or Partial, Dominance

Unlike the Mendelian crosses (reported in Chapter 3), a cross between parents with contrasting traits may sometimes generate offspring with an intermediate phenotype. For example, if a four-o'clock or a snapdragon plant with red flowers is crossed with a white-flowered plant, the offspring have pink flowers. Because some red pigment is produced in the F₁ intermediate-colored plant, neither the red nor white flower color is dominant. Such a situation is known as **incomplete, or partial, dominance**.

If the phenotype is under the control of a single gene and two alleles, where neither is dominant, the results of the F₁ (pink) × F₁ (pink) cross can be predicted. The resulting F₂ generation shown in Figure 4.1 confirms the hypothesis that only one pair of alleles determines these phenotypes. The genotypic ratio (1:2:1) of the F₂ generation is identical to that of Mendel's monohybrid cross. However, because neither allele is dominant, the phenotypic ratio is identical to the genotypic ratio (in contrast to the 3:1 phenotypic ratio of a Mendelian monohybrid cross). Note that because neither allele is recessive, we have chosen not to use upper- and lowercase letters as symbols. Instead, we denote the alleles responsible for red and white color as R¹ and R². We could have chosen W¹ and W² or still other designations such as C^W and C^R, where C indicates "color" and the W and R superscripts indicate "white" and "red," respectively.

How are we to interpret lack of dominance whereby an intermediate phenotype characterizes heterozygotes? The most accurate way is to consider gene expression in a quantitative way. In the case of flower color, the mutation causing

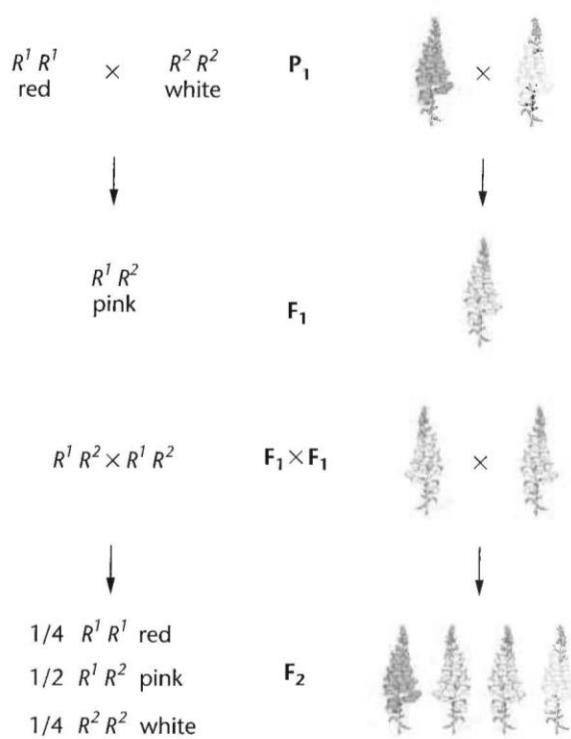


FIGURE 4.1 Incomplete dominance shown in the flower color of snapdragons.

white flowers is most likely one where complete “loss of function” occurs. In this case, it is likely that the gene product of the wild-type allele (R^1) is an enzyme that participates in a reaction leading to the synthesis of a red pigment. The mutant allele (R^2) produces an enzyme that cannot catalyze the reaction leading to pigment. The end result is that the heterozygote produces only about half the pigment of the red-flowered plant and the phenotype is pink.

Clear-cut cases of incomplete dominance are relatively rare. However, even when one allele seems to have complete dominance over the other, careful examination of the gene product and its activity, rather than the phenotype, often reveals an intermediate level of gene expression. An example is the human biochemical disorder **Tay–Sachs disease**, previously discussed in Chapter 3 (see p. 91), in which homozygous recessive individuals are severely affected with a fatal lipid-storage disorder and neonates

die during their first one to three years of life. In afflicted individuals, there is almost no activity of **hexosaminidase A**, an enzyme normally involved in lipid metabolism. Heterozygotes, with only a single copy of the mutant gene, are phenotypically normal, but with only about 50 percent of the enzyme activity found in homozygous normal individuals. Fortunately, this level of enzyme activity is adequate to achieve normal biochemical function. This situation is not uncommon in enzyme disorders and illustrates the concept of the **threshold effect**, whereby normal phenotypic expression occurs anytime a minimal level of gene product is attained. Most often, and in particular in Tay–Sachs disease, the threshold is less than 50 percent.

4.4 In Codominance, the Influence of Both Alleles in a Heterozygote Is Clearly Evident

If two alleles of a single gene are responsible for producing two distinct, detectable gene products, a situation different from incomplete dominance or dominance/recessiveness arises. In this case, *the joint expression of both alleles in a heterozygote* is called **codominance**. The **MN blood group** in humans illustrates this phenomenon. Karl Landsteiner and Philip Levine discovered a glycoprotein molecule found on the surface of red blood cells that acts as a native antigen, providing biochemical and immunological identity to individuals. In the human population, two forms of this glycoprotein exist, designated M and N; an individual may exhibit either one or both of them.

The MN system is under the control of a locus found on chromosome 4, with two alleles designated L^M and L^N . Because humans are diploid, three combinations are possible, each resulting in a distinct blood type:

Genotype	Phenotype
$L^M L^M$	M
$L^M L^N$	MN
$L^N L^N$	N

As predicted, a mating between two heterozygous MN parents may produce children of all three blood types, as follows:

$$\begin{array}{c} L^M L^N \times L^M L^N \\ \downarrow \\ 1/4 L^M L^M \\ 1/2 L^M L^N \\ 1/4 L^N L^N \end{array}$$

Once again, the genotypic ratio 1:2:1 is upheld.

Codominant inheritance is characterized by *distinct expression of the gene products of both alleles*. This characteristic distinguishes codominance from incomplete dominance, where heterozygotes express an intermediate, blended phenotype. For codominance to be studied, both products must be phenotypically detectable. We shall see another example of codominance when we examine the ABO blood-type system.

4.5 Multiple Alleles of a Gene May Exist in a Population

The information stored in any gene is extensive, and mutations can modify this information in many ways. Each change produces a different allele. Therefore, for any gene, the number of alleles within members of a population need not be restricted to two. When three or more alleles of the same gene—which we designate as **multiple alleles**—are present in a population, the resulting mode of inheritance may be unique. It is important to realize that *multiple alleles can be studied only in populations*. Any individual diploid organism has, at most, two homologous gene loci that may be occupied by different alleles of the same gene. However, among members of a species, numerous alternative forms of the same gene can exist.

The ABO Blood Groups

The simplest case of multiple alleles occurs when three alternative alleles of one gene exist. This situation is illustrated in the inheritance of the **ABO blood groups** in humans, discovered by Karl Landsteiner in the early 1900s. The ABO system, like the MN blood types, is characterized by the presence of antigens on the surface of red blood cells. The A and B antigens are distinct from the MN antigens and are under the control of a different gene, located on chromosome 9. As in the MN system, one combination of alleles in the ABO system exhibits a codominant mode of inheritance.

The ABO phenotype of any individual is ascertained by mixing a blood sample with an antiserum containing type A or type B antibodies. If an antigen is present on the surface of the person's red blood cells, it will react with the corresponding antibody and cause clumping, or agglutination, of the red blood cells. When an individual is tested in this way, one of four phenotypes may be revealed. Each individual has the A antigen (A phenotype), the B antigen (B phenotype), the A and B antigens (AB phenotype), or neither antigen (O phenotype).

In 1924, it was hypothesized that these phenotypes were inherited as the result of three alleles of a single gene. This hypothesis was based on studies of the blood types of many different families. Although different designations can be used, we will use the symbols I^A , I^B , and i to distinguish these three alleles. The I designation stands for *isoagglutinogen*, another term for antigen. If we assume that the I^A and I^B alleles are responsible for the production of their respective A and B antigens and that i is an allele that does not produce any detectable A or B antigens, we can list the various genotypic possibilities and assign the appropriate phenotype to each:

Genotype	Antigen	Phenotype
$I^A I^A$	A}	
$I^A i$	A}	A
$I^B I^B$	B}	
$I^B i$	B}	B
$I^A I^B$	A, B	AB
$i i$	Neither	O

In these assignments, the I^A and I^B alleles are dominant to the i allele, but codominant to each other.

Our knowledge of human blood types has several practical applications, including compatible blood transfusions and successful organ transplants.

The A and B Antigens

The biochemical basis of the ABO blood type system has now been carefully worked out. The A and B antigens are actually carbohydrate groups (sugars) that are bound to lipid molecules (fatty acids) protruding from the membrane of the red blood cell. The specificity of the A and B antigens is based on the terminal sugar of the carbohydrate group.

Almost all individuals possess what is called the **H substance**, to which one or two terminal sugars are added. As shown in Figure 4.2, the H substance itself contains three sugar molecules—galactose (Gal), *N*-acetylglucosamine (AcGluNH), and fucose—chemically linked together. The I^A allele is responsible for an enzyme that can add the terminal sugar *N*-acetylgalactosamine (AcGalNH) to the H substance. The I^B allele is responsible for a modified enzyme that cannot add *N*-acetylgalactosamine, but instead can add a terminal galactose. Heterozygotes ($I^A I^B$) add either one or the other sugar at the many sites (substrates) available on the surface of the red blood cell, illustrating the biochemical basis of codominance in individuals of the AB blood type. Finally, persons of type O ($i i$) cannot add either terminal sugar; these persons have only the H substance protruding from the surface of their red blood cells.

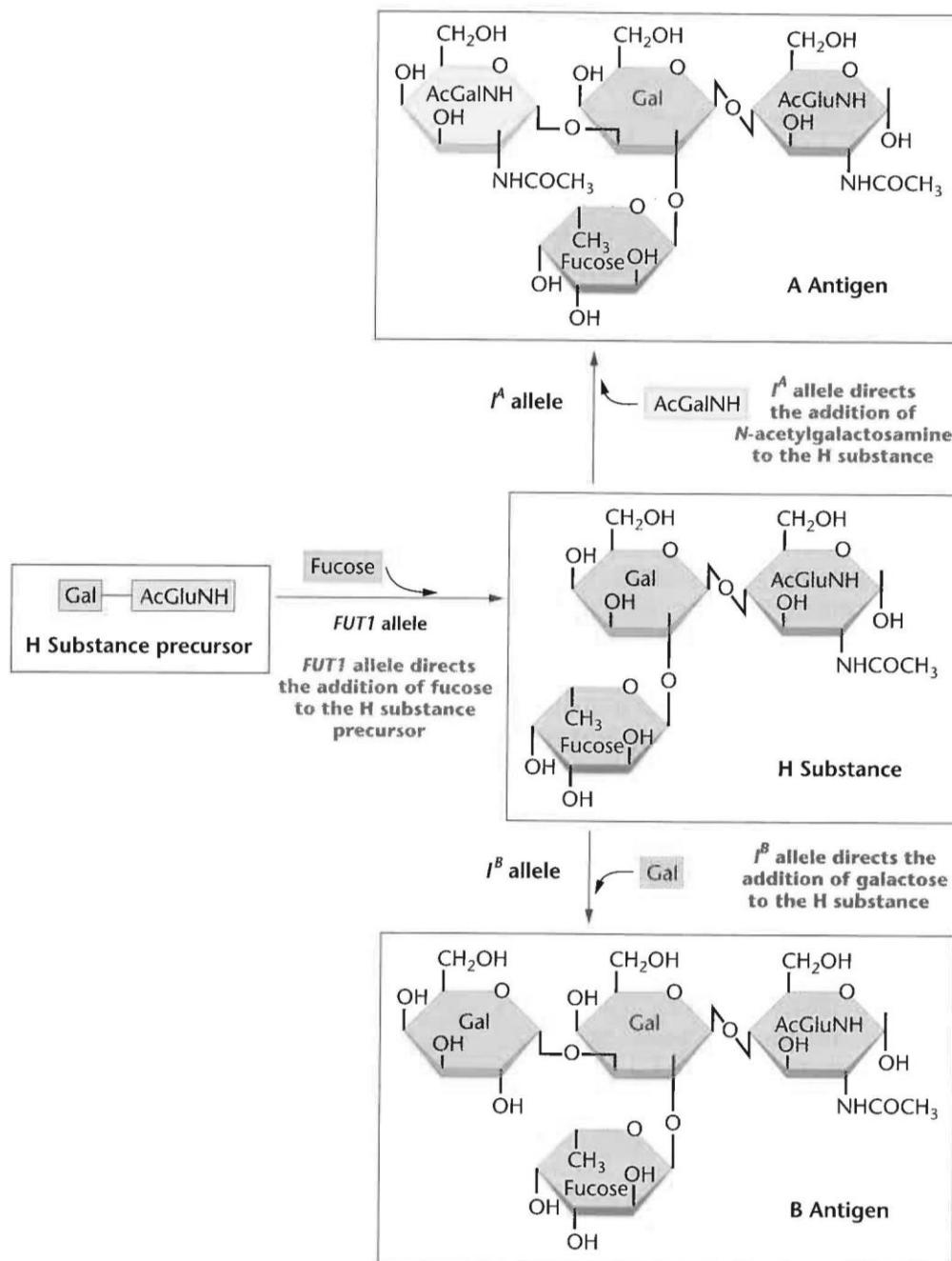


FIGURE 4.2 The biochemical basis of the ABO blood groups. The wild-type *FUT1* allele, present in almost all humans, directs the conversion of a precursor molecule to the H substance by adding a molecule of fucose to it. The *I^A* and *I^B* alleles are then able to direct the addition of terminal sugar residues to the H substance. The *i* allele is unable to

direct either of these terminal additions. Failure to produce the H substance results in the Bombay phenotype, in which individuals are type O regardless of the presence of an *I^A* or *I^B* allele. Gal: galactose; AcGluNH: *N*-acetylglycosamine; AcGalNH: *N*-acetylgalactosamine.

The Bombay Phenotype

In 1952, a very unusual situation provided information concerning the genetic basis of the H substance. A woman in Bombay displayed a unique genetic history inconsistent with her blood type. In need of a transfusion, she was found

to lack both the A and B antigens and was thus typed as O. However, as shown in the partial pedigree in Figure 4.3, one of her parents was type AB, and she herself was the obvious donor of an *I^B* allele to two of her offspring. Thus, she was genetically type B but functionally type O!

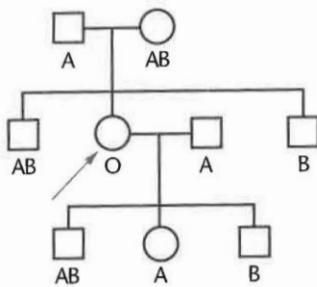


FIGURE 4.3 A partial pedigree of a woman with the Bombay phenotype. Functionally, her ABO blood group behaves as type O. Genetically, she is type B.

This woman was subsequently shown to be homozygous for a rare recessive mutation in a gene designated *FUT1* (encoding an enzyme, fucosyl transferase), which prevented her from synthesizing the complete H substance. In this mutation, the terminal portion of the carbohydrate chain protruding from the red cell membrane lacks fucose, normally added by the enzyme. In the absence of fucose, the enzymes specified by the I^A and I^B alleles apparently are unable to recognize the incomplete H substance as a proper substrate. Thus, neither the terminal galactose nor *N*-acetylgalactosamine can be added, even though the appropriate enzymes capable of doing so are present and functional. As a result, the ABO system genotype cannot be expressed in individuals homozygous for the mutant form of the *FUT1* gene; even though they may have the I^A and/or the I^B alleles, neither antigen is added to the cell surface, and they are functionally type O. To distinguish them from the rest of the population, they are said to demonstrate the **Bombay phenotype**. The frequency of the mutant *FUT1* allele is exceedingly low. Hence, the vast majority of the human population can synthesize the H substance.

The white Locus in *Drosophila*

Many other phenotypes in plants and animals are influenced by multiple allelic inheritance. In *Drosophila*, many alleles are present at practically every locus. The recessive mutation that causes white eyes, discovered by Thomas H. Morgan and Calvin Bridges in 1912, is one of over 100 alleles that can occupy this locus. In this allelic series, eye colors range from complete absence of pigment in the *white* allele to deep ruby in the *white-satsuma* allele, orange in the *white-apricot* allele, and a buff color in the *white-buff* allele. These alleles are designated w , w^{sat} , w^a , and w^{bf} , respectively. In each case, the total amount of pigment in these mutant eyes is reduced to less than 20 percent of that found in the brick-red wild-type eye. **Table 4.1** lists these and other *white* alleles and their color phenotypes.

It is interesting to note the biological basis of the original *white* mutation in *Drosophila*. Given what we know about eye color in this organism, it might be logical

TABLE 4.1 Some of the Alleles Present at the *white* Locus of *Drosophila*

Allele	Name	Eye Color
w	white	pure white
w^a	white-apricot	yellowish orange
w^{bf}	white-buff	light buff
w^{bl}	white-blood	yellowish ruby
w^{cf}	white-coffee	deep ruby
w^e	white-eosin	yellowish pink
w^{mo}	white-mottled orange	light mottled orange
w^{sat}	white-satsuma	deep ruby
w^{sp}	white-spotted	fine grain, yellow mottling
w^t	white-tinged	light pink

to presume that the mutant allele somehow interrupts the biochemical synthesis of pigments making up the brick-red eye of the wild-type fly. However, it is now clear that the product of the *white* locus is a protein that is involved in transporting pigments into the ommatidia (the individual units) comprising the compound eye. While flies expressing the *white* mutation can synthesize eye pigments normally, they cannot transport them into these structural units of the eye, thus rendering the white phenotype.

NOW SOLVE THIS

4.1 In the guinea pig, one locus involved in the control of coat color may be occupied by any of four alleles: C (full color), c^k (sepia), c^d (cream), or c^a (albino), with an order of dominance of: $C > c^k > c^d > c^a$. (C is dominant to all others, c^k is dominant to c^d and c^a , but not C , etc.) In the following crosses, determine the parental genotypes and predict the phenotypic ratios that would result:

- sepia \times cream, where both guinea pigs had an albino parent
- sepia \times cream, where the sepia guinea pig had an albino parent and the cream guinea pig had two sepia parents
- sepia \times cream, where the sepia guinea pig had two full-color parents and the cream guinea pig had two sepia parents
- sepia \times cream, where the sepia guinea pig had a full-color parent and an albino parent and the cream guinea pig had two full-color parents

■ **HINT:** This problem involves an understanding of multiple alleles. The key to its solution is to note particularly the hierarchy of dominance of the various alleles. Remember also that even though there can be more than two alleles in a population, an individual can have at most two of these. Thus, the allelic distribution into gametes adheres to the principle of segregation.

4.6 Lethal Alleles Represent Essential Genes

Many gene products are essential to an organism's survival. Mutations resulting in the synthesis of a gene product that is nonfunctional can often be tolerated in the heterozygous state; that is, one wild-type allele may be sufficient to produce enough of the essential product to allow survival. However, such a mutation behaves as a **recessive lethal allele**, and homozygous recessive individuals will not survive. The time of death will depend on when the product is essential. In mammals, for example, this might occur during development, early childhood, or even adulthood.

In some cases, the allele responsible for a lethal effect when homozygous may also result in a distinctive mutant phenotype when present heterozygously. *It is behaving as*

a recessive lethal allele but is dominant with respect to the phenotype. For example, a mutation that causes a yellow coat in mice was discovered in the early part of this century. The yellow coat varies from the normal agouti (wild-type) coat phenotype, as shown in [Figure 4.4](#). Crosses between the various combinations of the two strains yield unusual results:

Crosses			
(A) agouti	×	agouti	→ all agouti
(B) yellow	×	yellow	→ 2/3 yellow 1/3 agouti
(C) agouti	×	yellow	→ 1/2 yellow 1/2 agouti

These results are explained on the basis of a single pair of alleles. With regard to coat color, the mutant *yellow* allele (A^Y) is dominant to the wild-type *agouti* allele (A), so heterozygous mice will have yellow coats. However, the *yellow* allele is also a homozygous recessive lethal. When present

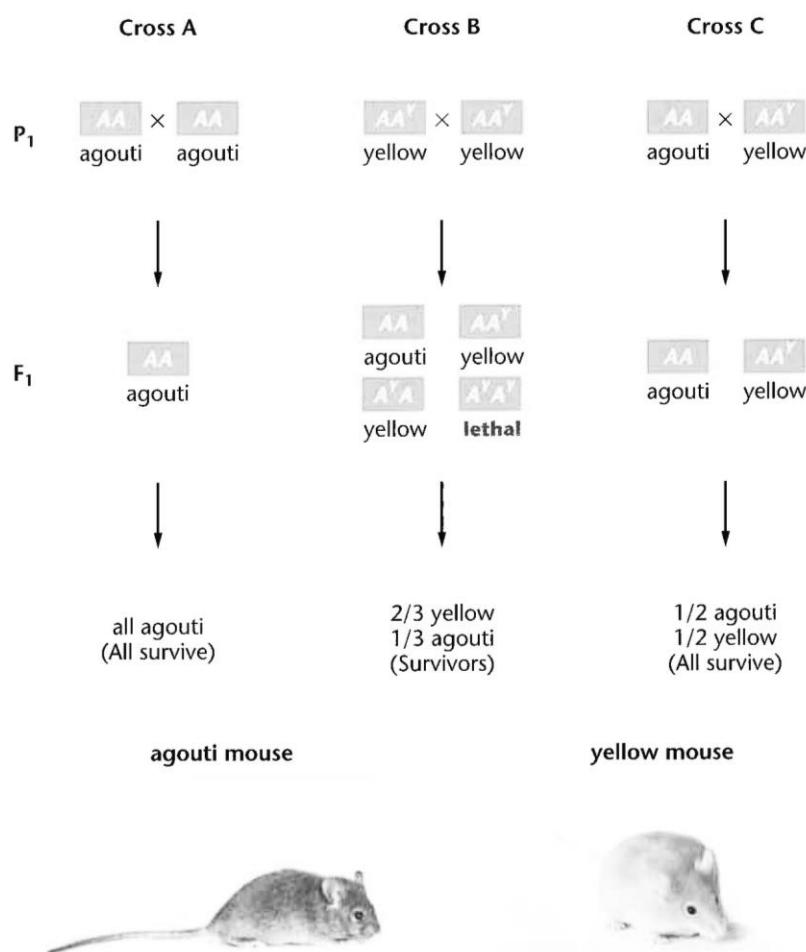


FIGURE 4.4 Inheritance patterns in three crosses involving the *agouti* allele (A) and the mutant *yellow* allele (A^Y) in the mouse. Note that the mutant allele behaves dominantly to the normal allele in controlling coat color, but it also behaves as a homozygous recessive lethal allele. Mice with the genotype $A^Y A^Y$ do not survive.

in two copies, the mice die before birth. Thus, there are no homozygous yellow mice. The genetic basis for these three crosses is shown in Figure 4.4.

In other cases, a mutation may behave as a **dominant lethal allele**. In such cases, the presence of just one copy of the allele results in the death of the individual. In humans, a disorder called **Huntington disease** is due to a dominant autosomal allele *H*, where the onset of the disease in heterozygotes (*Hh*) is delayed, usually well into adulthood. Affected individuals then undergo gradual nervous and motor degeneration until they die. This lethal disorder is particularly tragic because it has such a late onset, typically at about age 40. By that time, the affected individual may have produced a family, and each of their children has a 50 percent probability of inheriting the lethal allele, transmitting the allele to his or her offspring, and eventually developing the disorder. The American folk singer and composer Woody Guthrie (father of folk singer Arlo Guthrie) died from this disease at age 55.

Dominant lethal alleles are rarely observed. For these alleles to exist in a population, the affected individuals must reproduce before the lethal allele is expressed, as can occur in Huntington disease. If all affected individuals die before reaching reproductive age, the mutant gene will not be passed to future generations, and the mutation will disappear from the population unless it arises again as a result of a new mutation.

The Molecular Basis of Dominance, Recessiveness, and Lethality: The *agouti* Gene

Molecular analysis of the gene resulting in the agouti and yellow mice has provided insight into how a mutation can be both dominant for one phenotypic effect (hair color) and recessive for another (embryonic development). The *A^Y* allele is a classic example of a gain-of-function mutation. Animals homozygous for the wild-type *A* allele have yellow pigment deposited as a band on the otherwise black hair shaft, resulting in the agouti phenotype (see Figure 4.4). Heterozygotes deposit yellow pigment along the entire length of hair shafts as a result of the deletion of the regulatory region preceding the DNA coding region of the *A^Y* allele. Without any means to regulate its expression, one copy of the *A^Y* allele is always turned on in heterozygotes, resulting in the gain of function leading to the dominant effect.

The homozygous lethal effect has also been explained by molecular analysis of the mutant gene. The extensive deletion of genetic material that produced the *A^Y* allele actually extends into the coding region of an adjacent gene (*Merc*), rendering it nonfunctional. It is this gene that is critical to embryonic development, and the loss of its

function in *A^Y/A^Y* homozygotes is what causes lethality. Heterozygotes exceed the threshold level of the wild-type *Merc* gene product and thus survive.

1.2.4

Klug et al. (2019)

Page 110

Figure 4.8 + lecturer's comment

Case	Organism	Character	F ₂ Phenotypes				Modified ratio
			9/16	3/16	3/16	1/16	
1	Mouse	Coat color	agouti	albino	black	albino	9:3:4
2	Squash	Color	white		yellow	green	12:3:1
3	Pea	Flower color	purple	white			9:7
4	Squash	Fruit shape	disc	sphere		long	9:6:1
5	Chicken	Color	white		colored	white	13:3
6	Mouse	Color	white-spotted	white	colored	white-spotted	10:3:3
7	Shepherd's purse	Seed capsule	triangular			ovoid	15:1
8	Flour beetle	Color	6/16 sooty and 3/16 red	black	jet	black	6:3:3:4

FIGURE 4.8 The basis of modified dihybrid F₂ phenotypic ratios resulting from crosses between doubly heterozygous F₁ individuals. The four groupings of the F₂ genotypes shown in Figure 4.7 and across the top of this figure are combined in various ways to produce these ratios.

Comment lecturer

A double heterozygous cross looks like this:

P	AABB x aabb
F ₁	100 % AaBb
AaBb x AaBb	
F ₂	1/16 AABB
	2/16 AABb
	1/16 AAbb
	2/16 AaBB
	4/16 AaBb
	2/16 Aabb
	1/16 aaBB
	2/16 aaBb
	1/16 aabb

This type of cross can show us how two genes work together to create a specific phenotype (*epistasis*).

For example, if the genotype AA and Aa gives rise to no coloration, regardless of the genes at the B-locus, you will see a ratio of 12 : 4 for the phenotypes white : colored (case 2 in figure 4.8).

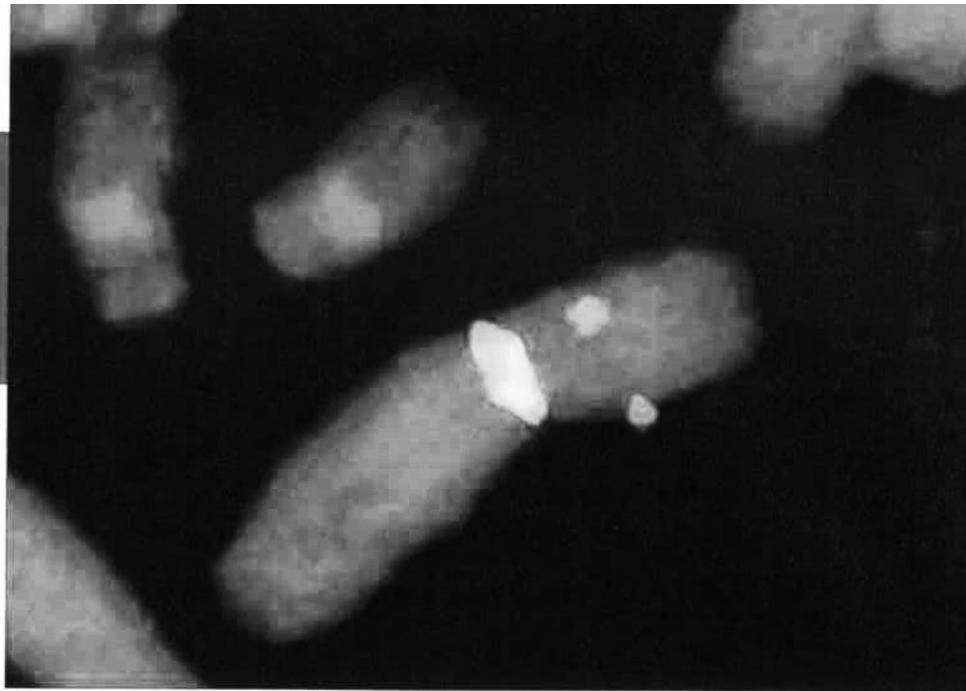
Challenge! Can you find out which genotypes belong to the phenotype groups for each case in figure 4.8?

1.2.5

Klug et al. (2019)

Page 131-144

5



Sex Determination and Sex Chromosomes

CHAPTER CONCEPTS

- A variety of mechanisms have evolved that result in sexual differentiation, leading to sexual dimorphism and greatly enhancing the production of genetic variation within species.
- Often, specific genes, usually on a single chromosome, cause maleness or femaleness during development.
- In humans, the presence of extra X or Y chromosomes beyond the diploid number may be tolerated but often leads to syndromes demonstrating distinctive phenotypes.
- While segregation of sex-determining chromosomes should theoretically lead to a one-to-one sex ratio of males to females, in humans the actual ratio favors males at conception.
- In mammals, females inherit two X chromosomes compared to one in males, but the extra genetic information in females is compensated for by random inactivation of one of the X chromosomes early in development.
- In some reptilian species, temperature during incubation of eggs determines the sex of offspring.

In the biological world, a wide range of reproductive modes and life cycles are observed. Some organisms are entirely asexual, displaying no evidence of sexual reproduction. Other organisms alternate between short periods of sexual reproduction and prolonged periods of

A human X chromosome highlighted using fluorescence *in situ* hybridization (FISH), a method in which specific probes bind to specific sequences of DNA. The green fluorescence probe binds to DNA at the centromere of X chromosomes. The red fluorescence probe binds to the DNA sequence of the X-linked Duchenne muscular dystrophy (DMD) gene.

asexual reproduction. In most diploid eukaryotes, however, sexual reproduction is the only natural mechanism for producing new members of the species. The perpetuation of all sexually reproducing organisms depends ultimately on an efficient union of gametes during fertilization. In turn, successful fertilization depends on some form of **sexual differentiation** in the reproductive organisms. Even though it is not overtly evident, this differentiation occurs in organisms as low on the evolutionary scale as bacteria and single-celled eukaryotic algae. In more complex forms of life, the differentiation of the sexes is more evident as phenotypic dimorphism of males and females. The ancient symbol for iron and for Mars, depicting a shield and spear (δ), and the ancient symbol for copper and for Venus, depicting a mirror (φ), have also come to symbolize maleness and femaleness, respectively.

Dissimilar, or **heteromorphic, chromosomes**, such as the XY pair in mammals, characterize one sex or the other in a wide range of species, resulting in their label as **sex chromosomes**. Nevertheless, it is genes, rather than chromosomes, that ultimately serve as the underlying basis of **sex determination**. As we will see, some of these genes are present on sex chromosomes, but others are autosomal. Extensive investigation has revealed a wide variation in sex-chromosome systems—even in closely related organisms—suggesting that mechanisms controlling sex

determination have undergone rapid evolution many times in the history of life.

In this chapter, we delve more deeply into what is known about the genetic basis for the determination of sexual differences, with a particular emphasis on two organisms: our own species, representative of mammals; and *Drosophila*, on which pioneering sex-determining studies were performed.

5.1 X and Y Chromosomes Were First Linked to Sex Determination Early in the Twentieth Century

How sex is determined has long intrigued geneticists. In 1891, Hermann Henking identified a nuclear structure in the sperm of certain insects, which he labeled the X-body. Several years later, Clarence McClung showed that some of the sperm in grasshoppers contain an unusual genetic structure, called a *heterochromosome*, but the remainder of the sperm lack this structure. He mistakenly associated the presence of the heterochromosome with the production of male progeny. In 1906, Edmund B. Wilson clarified Henking and McClung's findings when he demonstrated that female somatic cells in the butterfly *Protenor* contain 14 chromosomes, including 12 autosomes (A) and two X chromosomes. During oogenesis, an even reduction occurs, producing gametes with seven chromosomes, including one X chromosome ($6A + X$). Male somatic cells, on the other hand, contain only 13 chromosomes, including one X chromosome. During spermatogenesis, gametes are produced containing either six chromosomes, without an X ($6A$), or seven chromosomes, one of which is an X ($6A + X$). Fertilization by X-bearing sperm results in female offspring, and fertilization by X-deficient sperm results in male offspring [Figure 5.1(a)].

The presence or absence of the X chromosome in male gametes provides an efficient mechanism for sex determination in this species and also produces a 1:1 sex ratio in the resulting offspring.

Wilson also experimented with the milkweed bug *Lygaeus turcicus*, in which both sexes have 14 chromosomes. Twelve of these are autosomes. In addition, the females have two X chromosomes, while the males have only a single X and a smaller heterochromosome labeled the **Y chromosome**. Females in this species produce only gametes of the ($6A + X$) constitution, but males produce two types of gametes in equal proportions, ($6A + X$) and ($6A + Y$). Therefore, following random fertilization, equal numbers of male and female progeny will be produced with distinct chromosome complements [Figure 5.1(b)].

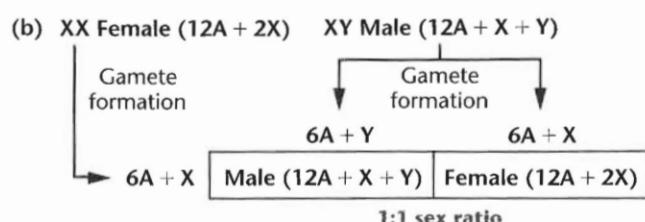
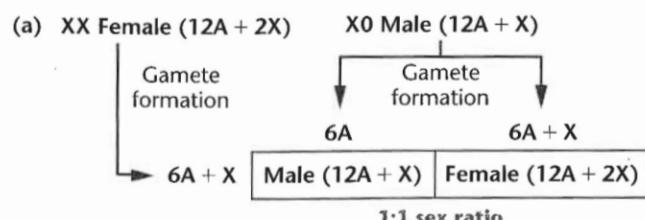


FIGURE 5.1 (a) Sex determination where the heterogametic sex (the male in this example) is X0 and produces gametes with or without the X chromosome; (b) sex determination, where the heterogametic sex (again, the male in this example) is XY and produces gametes with either an X or a Y chromosome. In both cases, the chromosome composition of the offspring determines its sex.

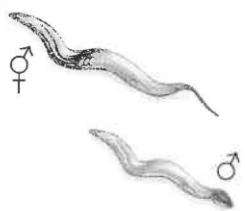
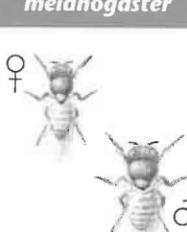
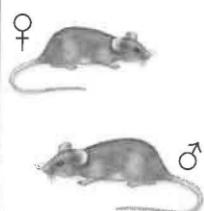
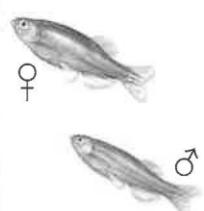
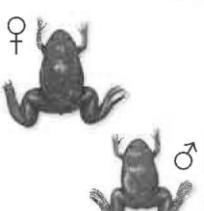
In *Protenor* and *Lygaeus*, males produce gametes with different chromosome compositions. As a result, they are described as the **heterogametic sex**, and in effect, their gametes ultimately determine the sex of the progeny in those species. In such cases, the female, which has like sex chromosomes, is the **homogametic sex**, producing uniform gametes with regard to chromosome numbers and types.

The male is not always the heterogametic sex. In some organisms, the female produces unlike gametes, exhibiting either the *Protenor* XX/XO or *Lygaeus* XX/XY mode of sex determination. Examples include certain moths and butterflies, some fish, reptiles, amphibians, at least one species of plants (*Fragaria orientalis*), and most birds. To immediately distinguish situations in which the female is the heterogametic sex, some geneticists use the notation ZZ/ZW, where ZZ is the homogametic male and ZW is the heterogametic female, instead of the XX/XY notation. For example, chickens are so denoted. The sex chromosome composition for popular model organisms in genetics is shown in Table 5.1.

5.2 The Y Chromosome Determines Maleness in Humans

The first attempt to understand sex determination in our own species occurred almost 100 years ago and involved the visual examination of chromosomes in dividing cells.

TABLE 5.1 Sex Chromosome Compositions of Common Model Organisms

	<i>Caenorhabditis elegans</i>	<i>Drosophila melanogaster</i>	<i>Mus musculus</i>	<i>Danio rerio</i>	<i>Xenopus laevis</i>	
Model Organism						
Sex Chromosomes	XX	XO	XX XY	XX XY	None	ZW ZZ

Efforts were made to accurately determine the diploid chromosome number of humans, but because of the relatively large number of chromosomes, this proved to be quite difficult. Then, in 1956, Joe Hin Tjio and Albert Levan discovered an effective way to prepare chromosomes for accurate viewing. This technique led to a strikingly clear demonstration of metaphase stages showing that 46 was indeed the human diploid number. Later that same year, C. E. Ford and John L. Hamerton, also working with testicular tissue, confirmed this finding. The familiar karyotypes of a human male (Figure 2.4) illustrate the difference in size between the human X and Y chromosomes.

Of the normal 23 pairs of human chromosomes, one pair was shown to vary in configuration in males and females. These two chromosomes were designated the X and Y sex chromosomes. The human female has two X chromosomes, and the human male has one X and one Y chromosome.

We might believe that this observation is sufficient to conclude that the Y chromosome determines maleness. However, several other interpretations are possible. The Y could play no role in sex determination; the presence of two X chromosomes could cause femaleness; or maleness could result from the lack of a second X chromosome. The evidence that clarified which explanation was correct came from study of the effects of human sex-chromosome variations, described in the following section. As such investigations revealed, the Y chromosome does indeed determine maleness in humans.

Klinefelter and Turner Syndromes

Around 1940, scientists identified two human abnormalities characterized by aberrant sexual development, **Klinefelter syndrome (47,XXY)** and **Turner syndrome (45,X)**.^{*} Individuals with Klinefelter syndrome are generally tall and have long arms and legs and large hands and feet. They usually have genitalia and internal ducts

that are male, but their testes are rudimentary and fail to produce sperm. At the same time, feminine sexual development is not entirely suppressed. Slight enlargement of the breasts (gynecomastia) is common, and the hips are often rounded. This ambiguous sexual development, referred to as intersexuality, can lead to abnormal social development. Intelligence is often below the normal range as well.

In Turner syndrome, the affected individual has female external genitalia and internal ducts, but the ovaries are rudimentary. Other characteristic abnormalities include short stature (usually under 5 feet), skin flaps on the back of the neck, and underdeveloped breasts. A broad, shieldlike chest is sometimes noted. Intelligence is usually normal.

In 1959, the karyotypes of individuals with these syndromes were determined to be abnormal with respect to the sex chromosomes. Individuals with Klinefelter syndrome have more than one X chromosome. Most often they have an XXY complement in addition to 44 autosomes [Figure 5.2(a)], which is why people with this karyotype are designated 47,XXY. Individuals with Turner syndrome most often have only 45 chromosomes, including just a single X chromosome; thus, they are designated 45,X [Figure 5.2(b)]. Note the convention used in designating these chromosome compositions. The number states the total number of chromosomes present, and the information after the comma indicates the deviation from the normal diploid content. Both conditions result from **nondisjunction**, the failure of the sex chromosomes to segregate properly during meiosis (nondisjunction is described in Chapter 6 and illustrated in Figure 6.1).

* Although the possessive form of the names of eponymous syndromes is sometimes used (e.g., Klinefelter's syndrome), the current preference is to use the nonpossessive form.

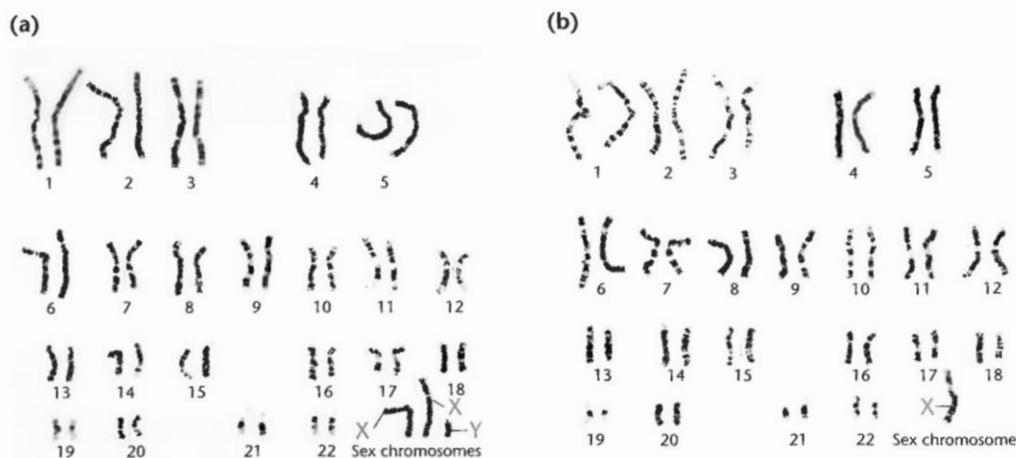


FIGURE 5.2 The karyotypes of individuals with (a) Klinefelter syndrome (47,XXY) and (b) Turner syndrome (45,X).

These Klinefelter and Turner karyotypes and their corresponding sexual phenotypes led scientists to conclude that the Y chromosome determines maleness in humans. In its absence, the person's sex is female, even if only a single X chromosome is present. The presence of the Y chromosome in the individual with Klinefelter syndrome is sufficient to determine maleness, even though male development is not complete. Similarly, in the absence of a Y chromosome, as in the case of individuals with Turner syndrome, no masculinization occurs. Note that we cannot conclude anything regarding sex determination under circumstances where a Y chromosome is present without an X because Y-containing human embryos lacking an X chromosome (designated 45,Y) do not survive.

Klinefelter syndrome occurs in about 1 of every 660 male births and is the most common sex chromosome disorder in males. The karyotypes **48,XXX**, **48,XXYY**, **49,XXXXY**, and **49,XXYY** are similar phenotypically to 47,XXY, but manifestations are often more severe in individuals with a greater number of X chromosomes. Recent studies have also shown that the variability in phenotypes for men with a 47,XXY genotype is correlated with copy number variations (CNVs), particularly duplications, on the X chromosomes.

Turner syndrome can also result from karyotypes other than 45,X, including individuals called **mosaics**, whose somatic cells display two different genetic cell lines, each exhibiting a different karyotype. Such cell lines result from a mitotic error during early development, the most common chromosome combinations being **45,X/46,XY** and **45,X/46,XX**. Thus, an embryo that began life with a normal karyotype can give rise to an individual whose cells show a mixture of karyotypes and who exhibits varying aspects of this syndrome.

Turner syndrome is observed in about 1 in 2000 female births, a frequency much lower than that for Klinefelter syndrome. One explanation for this difference is the observation that a substantial majority of 45,X fetuses die *in utero* and are aborted spontaneously. Thus, a similar frequency of the two syndromes may occur at conception.

47,XXX Syndrome

The abnormal presence of three X chromosomes along with a normal set of autosomes (**47,XXX**) results in female differentiation. The highly variable syndrome that accompanies this genotype, often called **triplo-X**, occurs in about 1 of 1000 female births. Frequently, 47,XXX women are perfectly normal and may remain unaware of their abnormality in chromosome number unless a karyotype is done. In other cases, underdeveloped secondary sex characteristics, sterility, delayed development of language and motor skills, and mental retardation may occur. In rare instances, **48,XXXX** (tetra-X) and **49,XXXXX** (penta-X) karyotypes have been reported. The syndromes associated with these karyotypes are similar to but more pronounced than the 47,XXX syndrome. Thus, in many cases, the presence of additional X chromosomes appears to disrupt the delicate balance of genetic information essential to normal female development.

47,XYY Condition

Another human condition involving the sex chromosomes is **47,XYY**. Studies of this condition, in which the only deviation from diploidy is the presence of an additional Y chromosome in an otherwise normal male karyotype, were initiated in 1965 by Patricia Jacobs. She discovered that 9 of 315 males in a Scottish maximum security prison had the 47,XYY karyotype. These males were significantly above average in height and had been incarcerated as a

result of dangerous, violent, or criminal propensities. Of the nine males studied, seven were of subnormal intelligence, and all suffered personality disorders. Several other studies produced similar findings.

The possible correlation between this chromosome composition and criminal behavior piqued considerable interest, and extensive investigation of the phenotype and frequency of the 47,XYY condition in both criminal and noncriminal populations ensued. Above-average height (usually over 6 feet) and subnormal intelligence were substantiated, and the frequency of males displaying this karyotype was indeed revealed to be higher in penal and mental institutions compared with unincarcerated populations [one study showed 29 XYY males when 28,366 were examined (0.10%)]. A particularly relevant question involves the characteristics displayed by the XYY males who are not incarcerated. The only nearly constant association is that such individuals are over 6 feet tall.

A study to further address this issue was initiated in 1974 to identify 47,XYY individuals at birth and to follow their behavioral patterns during preadult and adult development. While the study was considered unethical and soon abandoned, it has become clear that there are many XYY males present in the population who do not exhibit antisocial behavior and who lead normal lives. Therefore, we must conclude that there is a high, but not constant, correlation between the extra Y chromosome and the predisposition of these males to exhibit behavioral problems.

Sexual Differentiation in Humans

Once researchers had established that, in humans, it is the Y chromosome that houses genetic information necessary for maleness, they attempted to pinpoint a specific gene or genes capable of providing the “signal” responsible for sex determination. Before we delve into this topic, it is useful to consider how sexual differentiation occurs in order to better comprehend how humans develop into sexually dimorphic males and females. During early development, every human embryo undergoes a period when it is potentially hermaphroditic. By the fifth week of gestation, gonadal primordia (the tissues that will form the gonad) arise as a pair of **gonadal (genital) ridges** associated with each embryonic kidney. The embryo is potentially hermaphroditic because at this stage its gonadal phenotype is sexually indifferent—male or female reproductive structures cannot be distinguished, and the gonadal ridge tissue can develop to form male or female gonads. As development progresses, primordial germ cells migrate to these ridges, where an outer cortex and inner medulla form (*cortex* and *medulla* are the outer and inner tissues of an organ, respectively). The cortex is capable of developing into an ovary, while the medulla may develop into a testis. In addition,

two sets of undifferentiated ducts called the Wolffian and Müllerian ducts exist in each embryo. Wolffian ducts differentiate into other organs of the male reproductive tract, while Müllerian ducts differentiate into structures of the female reproductive tract.

Because gonadal ridges can form either ovaries or testes, they are commonly referred to as **bipotential gonads**. What switch triggers gonadal ridge development into testes or ovaries? The presence or absence of a Y chromosome is the key. If cells of the ridge have an XY constitution, development of the medulla into a testis is initiated around the seventh week. However, in the absence of the Y chromosome, no male development occurs, the cortex of the ridge subsequently forms ovarian tissue, and the Müllerian duct forms oviducts (Fallopian tubes), uterus, cervix, and portions of the vagina. Depending on which pathway is initiated, parallel development of the appropriate male or female duct system then occurs, and the other duct system degenerates. If testes differentiation is initiated, the embryonic testicular tissue secretes hormones that are essential for continued male sexual differentiation. As we will discuss in the next section, the presence of a Y chromosome and the development of the testes also inhibit formation of female reproductive organs.

In females, as the twelfth week of fetal development approaches, the oogonia within the ovaries begin meiosis, and primary oocytes can be detected. By the twenty-fifth week of gestation, all oocytes become arrested in meiosis and remain dormant until puberty is reached some 10 to 15 years later. In males, on the other hand, primary spermatocytes are not produced until puberty is reached (see Figure 2.11).

As sexual dimorphism is considered, it is important to distinguish between *primary sexual differentiation*, which involves only the gonads, where gametes are produced, and *secondary sexual differentiation*, which involves the overall phenotype of the organism. Secondary effects include clear differences in such organs as mammary glands and external genitalia as well as other traits that differ between males and females.

The Y Chromosome and Male Development

The human Y chromosome, unlike the X, was long thought to be mostly blank genetically. It is now known that this is not true, even though the Y chromosome contains far fewer genes than does the X. Data from the Human Genome Project indicate that the Y chromosome has at least 75 genes, compared to 900–1400 genes on the X. Current analysis of these genes and regions with potential genetic function reveals that some have homologous counterparts on the X chromosome and others do not. In addition, recent work has revealed that a small number of conserved and essential genes previously thought to be lost from the Y chromosome throughout evolution are present on autosomes. Present on both ends of

the Y chromosome are so-called **pseudoautosomal regions (PARs)** that share homology with regions on the X chromosome and synapse and recombine with it during meiosis. The presence of such a pairing region is critical to segregation of the X and Y chromosomes during male gametogenesis. The remainder of the chromosome, about 95 percent of it, does not synapse or recombine with the X chromosome. As a result, it was originally referred to as the *nonrecombinant region of the Y (NRY)*. More recently, researchers have designated this region as the **male-specific region of the Y (MSY)**. Some portions of the MSY share homology with genes on the X chromosome, and others do not.

The human Y chromosome is diagrammed in Figure 5.3. The MSY is divided about equally between *euchromatic* regions, containing functional genes, and *heterochromatic* regions, lacking genes. Within euchromatin, adjacent to the PAR of the short arm of the Y chromosome, is a critical gene that controls male sexual development, called the **sex-determining region Y (SRY)**. In humans, the absence of a Y chromosome almost always leads to female development; thus, this gene is absent from the X chromosome. At six to eight weeks of development, the SRY gene becomes active in XY embryos. SRY encodes a protein that causes the undifferentiated gonadal tissue of the embryo to form testes. This protein is called the **testis-determining factor (TDF)**. SRY (or a closely related version) is present in all mammals thus far examined, indicative of its essential function throughout this diverse group of animals.*

Our ability to identify the presence or absence of DNA sequences in rare individuals whose sex-chromosome composition does not correspond to their sexual phenotype has provided evidence that SRY is the gene responsible for male sex determination. For example, there are human males

who have two X and no Y chromosomes. Often, attached to one of their X chromosomes is the region of the Y that contains SRY. There are also females who have one X and one Y chromosome, a condition known as XY sex reversal or Swyer syndrome. Their Y is almost always missing the SRY gene or they have a specific mutation in SRY. These observations argue strongly in favor of the role of SRY in providing the primary signal for male development.

Further support of this conclusion involves an experiment using **transgenic mice**. These animals are produced from fertilized eggs injected with foreign DNA that is subsequently incorporated into the genetic composition of the developing embryo. In normal mice, a chromosome region designated *Sry* has been identified that is comparable to SRY in humans. When mouse DNA containing *Sry* is injected into normal XX mouse eggs, most of the offspring develop into males.

The question of how the product of this gene triggers development of embryonic gonadal tissue into testes rather than ovaries has been under investigation for 25 years. TDF functions as a *transcription factor*, a DNA-binding protein that interacts directly with regulatory sequences of other genes to stimulate their expression. Thus, while TDF behaves as a master switch that controls other genes downstream in the process of sexual differentiation, identifying TDF target genes has been difficult. One potential target for activation by TDF that has been extensively studied is the gene for **Müllerian inhibiting substance (MIS)**, [also called Mullerian inhibiting hormone, (MIH), or anti-Mullerian hormone]. Cells of the developing testes secrete MIS. As its name suggests, MIS protein causes regression (atrophy) of cells in the Müllerian duct. Degeneration of the duct prevents formation of the female reproductive tract.

Other autosomal genes are part of a cascade of genetic expression initiated by SRY. Examples include the human *SOX9* gene and the mouse homolog *Sox9*, which when activated by SRY, leads to the differentiation of cells that form the seminiferous tubules that contain male germ cells. In the mouse, fibroblast growth factor 9 (*Fgf9*) is upregulated in XY gonads. Testis development is completely blocked in gonads lacking *Fgf9*, and signs of ovarian development occur. Another gene, *SFI*, is involved in the regulation of enzymes affecting steroid metabolism. In mice, this gene is initially active in both the male and female bisexual genital ridge, persisting until

the point in development when testis formation is apparent. At that time, its expression persists in males but is extinguished in females. Recent work using mice has

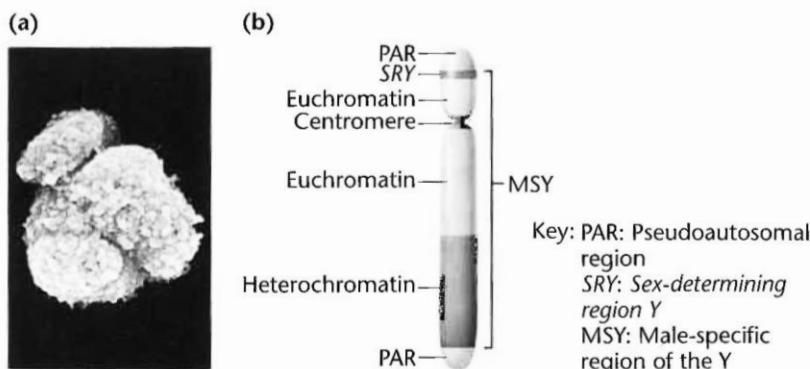


FIGURE 5.3 (a) Electron micrograph of the human Y chromosome (magnification $\times 35,000$) and (b) regions of the Y chromosome.

* It is interesting to note that in chickens, a similar gene has recently been identified. Called *DMRT1*, it is located on the Z chromosome. This gene is the subject of Problem 29 in the Problems section at the end of the chapter.

suggested that testicular development may be actively repressed throughout the life of females by downregulating expression of specific genes. This is based on experiments showing that, in adult female mice, deletion of a gene *Foxl2*, which encodes a transcription factor, leads to transdifferentiation of the ovary into the testis.

In 2016, researchers at the University of Hawaii published novel work demonstrating that two genes in mice, *Sox9* and *Eif2s3y*, could substitute for the Y chromosome. *Sry* activates *Sox9*, and *Eif2s3y* has a homolog on the X chromosome (*Eif2s3x*). Transgenic mice with one X and no Y chromosome were generated. But in these mice, *Sry* was replaced with a transgenic copy of *Sox9* and made to overexpress *Eif2s3x* from an X chromosome, beyond the levels produced normally by the X and Y chromosomes. These males, lacking a Y chromosome, produced haploid male gametes. They did not produce mature sperm but yielded round spermatids that were used to fertilize an oocyte *in vitro*, resulting in viable offspring. This study demonstrated that *Sox9*, in the absence of *Sry*, and the *Eif2s3y* homolog, *Eif2s3x*, allow for male gamete development and initiation of spermatogenesis in the absence of a complete Y chromosome. While these two genes can result in male gametes that produce offspring through assisted reproductive technology, other genes are necessary to produce mature sperm, but nonetheless, experiments such as these are providing novel insights into the genetics of sex-determination pathways. Establishment of the link between these various genes and sex determination has brought us closer to a complete understanding of how males and females arise in humans, but much work remains to be done.

Findings by David Page and his many colleagues have provided a reasonably complete picture of the MSY region of the human Y chromosome. Page has spearheaded the detailed study of the Y chromosome for the past several decades. The MSY consists of about 23 million base pairs (23 Mb) and can be divided into three regions. The first region is the *X-transposed region*. It comprises about 15 percent of the MSY and was originally derived from the X chromosome in the course of human evolution (about 3 to 4 million years ago). The X-transposed region is 99 percent identical to region Xq21 of the modern human X chromosome. Two genes, both with X chromosome homologs, are present in this region.

Research by Page and others has also revealed that sequences called **palindromes**—sequences of base pairs that read the same but in the opposite direction on complementary strands—are present throughout the MSY. Recombination between palindromes on sister chromatids of the Y chromosome during replication is a mechanism

used to repair mutations in the Y chromosome. This discovery has fascinating implications concerning how the Y chromosome may maintain its size and structure.

Another interesting finding is that the MSY of the human Y chromosome is very different in sequence structure than the MSY from chimpanzees. The study indicates that rapid evolution has occurred since separation of these species over 6 million years ago—a surprise given that primate sex chromosomes have been in existence for hundreds of millions of years. Over 30 percent of the chimpanzee MSY sequence has no homologous sequence in the human MSY. The chimpanzee MSY has lost many protein-coding genes compared to common ancestors but contains twice the number of palindromic sequences as the human MSY.

The second area of the MSY is designated the *X-degenerative region*. Comprising about 20 percent of the MSY, this region contains DNA sequences that are even more distantly related to those present on the X chromosome. The X-degenerative region contains 27 single-copy genes and a number of *pseudogenes* (genes whose sequences have degenerated sufficiently during evolution to render them nonfunctional). Twenty of the 27 genes located here share homology with counterparts on the X chromosome and evolved from genes on the X chromosome. One of these is the *SRY* gene, discussed earlier. Other X-degenerative genes that encode protein products are expressed ubiquitously in all tissues in the body, but *SRY* is expressed only in the testes.

The third area, the *ampliconic region*, contains about 30 percent of the MSY, including most of the genes closely associated with the development of testes. These genes lack counterparts on the X chromosome, and their expression is limited to the testes. There are 60 transcription units (genes that yield a product) divided among nine gene families in this region, most represented by multiple copies. Members of each family have nearly identical (>98 percent) DNA sequences. Each repeat unit is an **amplicon** and is contained within seven segments scattered across the euchromatic regions of both the short and long arms of the Y chromosome. Genes in the ampliconic region encode proteins specific to the development and function of the testes, and the products of many of these genes are directly related to fertility in males. It is currently believed that a great deal of male sterility in our population can be linked to mutations in these genes.

Until relatively recently it was thought that the Y chromosome only contributed to sex determination and male fertility. A recent area of investigation has involved the Y chromosome and paternal age. For many years, it has been known that maternal age is correlated with an elevated rate of offspring with chromosomal defects, including Down syndrome (see Chapter 6). Advanced

paternal age has now been associated with an increased risk in offspring of congenital disorders with a genetic basis, including certain cancers, schizophrenia, autism, and other conditions, collectively known as *paternal age effects* (PAE). Studies in which the genomes of sperm have been sequenced have demonstrated the presence of specific PAE mutations including numerous ones on the Y chromosome. Evidence suggests that PAE mutations are positively selected for and result in an enrichment of mutant sperm over time.

Similarly, an analysis of blood samples and medical records for more than 6000 men in Sweden revealed a correlation between smoking and complete loss of the Y chromosome in blood cells. Y chromosome loss was also correlated to elevated cancer risk among male smokers, reduced expression of tumor-suppressor genes, and compromised immunity. This and other research provides further evidence that genes on the Y chromosome affect more than sex determination and male fertility.

This recent work has greatly expanded our picture of the genetic information carried by this unique chromosome. It clearly refutes the so-called *wasteland theory*, prevalent some 25 years ago, that depicted the human Y chromosome as almost devoid of genetic information other than a few genes that cause maleness. The knowledge we have gained provides the basis for a much clearer picture of how maleness is determined.

NOW SOLVE THIS

5.1 Campomelic dysplasia (CMD1) is a congenital human syndrome featuring malformation of bone and cartilage. It is caused by an autosomal dominant mutation of a gene located on chromosome 17. Consider the following observations in sequence, and in each case, draw whatever appropriate conclusions are warranted.

- Of those with the syndrome who are karyotypically 46,XY, approximately 75 percent are sex reversed, exhibiting a wide range of female characteristics.
- The nonmutant form of the gene, called SOX9, is expressed in the developing gonad of the XY male, but not the XX female.
- The SOX9 gene shares 71 percent amino acid coding sequence homology with the Y-linked SRY gene.
- CMD1 patients who exhibit a 46,XX karyotype develop as females, with no gonadal abnormalities.

■ **HINT:** This problem asks you to apply the information presented in this chapter to a real-life example. The key to its solution is knowing that some genes are activated and produce their normal product as a result of expression of products of other genes found on different chromosomes.

5.3 The Ratio of Males to Females in Humans Is Not 1.0

The presence of heteromorphic sex chromosomes in one sex of a species but not the other provides a potential mechanism for producing equal proportions of male and female offspring. This potential depends on the segregation of the X and Y (or Z and W) chromosomes during meiosis, such that half of the gametes of the heterogametic sex receive one of the chromosomes and half receive the other one. As we learned in Section 5.2, small pseudoautosomal regions of pairing homology do exist at both ends of the human X and Y chromosomes, suggesting that the X and Y chromosomes do synapse and then segregate into different gametes. Provided that both types of gametes are equally successful in fertilization and that the two sexes are equally viable during development, a 1:1 ratio of male and female offspring should result.

The actual proportion of male to female offspring, referred to as the **sex ratio**, has been assessed in two ways. The **primary sex ratio (PSR)** reflects the proportion of males to females conceived in a population. The **secondary sex ratio** reflects the proportion of each sex that is born. The secondary sex ratio is much easier to determine but has the disadvantage of not accounting for any disproportionate embryonic or fetal mortality.

When the secondary sex ratio in the human population was determined in 1969 by using worldwide census data, it did not equal 1.0. For example, in the Caucasian population in the United States, the secondary ratio was a little less than 1.06, indicating that about 106 males were born for each 100 females. (In 1995, this ratio dropped to slightly less than 1.05.) In the African-American population in the United States, the ratio was 1.025. In other countries, the excess of male births is even greater than is reflected in these values. For example, in Korea, the secondary sex ratio was 1.15.

Despite these ratios, it is possible that the PSR is 1.0 and is altered between conception and birth. For the secondary ratio to exceed 1.0, then, prenatal female mortality would have to be greater than prenatal male mortality. However, when this hypothesis was first examined, it was deemed to be false. In a Carnegie Institute study, reported in 1948, the sex of approximately 6000 embryos and fetuses recovered from miscarriages and abortions was determined, and fetal mortality was actually higher in males. On the basis of the data derived from that study, the PSR in U.S. Caucasians was estimated to be 1.079, suggesting that more males than females are conceived in the human population.

To explain why, researchers considered the assumptions on which the theoretical ratio is based:

1. Because of segregation, males produce equal numbers of X- and Y-bearing sperm.
2. Each type of sperm has equivalent viability and motility in the female reproductive tract.
3. The egg surface is equally receptive to both X- and Y-bearing sperm.

No direct experimental evidence contradicts any of these assumptions.

A PSR favoring male conceptions remained dogma for many decades until, in 2015, a study using an extensive dataset was published determining that the PSR is indeed 1.0, thus concluding that equal numbers of males and females are conceived. Among other parameters, the examination of the sex of 3-day-old and 6-day-old embryos conceived using assisted reproductive technology provided the most direct assessment. Following conception, however, mortality was then shown to fluctuate between the sexes, until at birth, more males than females are born. Thus, female mortality during embryonic and fetal development exceeds that of males. Clearly, this is a difficult topic to investigate but one of continued interest. For now, the most recent findings are convincing and contradict the earlier studies.

5.4 Dosage Compensation Prevents Excessive Expression of X-Linked Genes in Humans and Other Mammals

The presence of two X chromosomes in normal human females and only one X in normal human males is unique compared with the equal numbers of autosomes present in the cells of both sexes. On theoretical grounds alone, it is possible to speculate that this disparity should create a “genetic dosage” difference between males and females, with attendant problems, for all X-linked genes. There is the potential for females to produce twice as much of each product of all X-linked genes. The additional X chromosomes in both males and females exhibiting the various syndromes discussed earlier in this chapter are thought to compound this dosage problem. Embryonic development depends on proper timing and precisely regulated levels of gene expression. Otherwise, disease phenotypes or embryonic lethality can occur. In this section, we will describe research findings regarding X-linked gene expression that demonstrate a genetic mechanism of **dosage compensation** that balances the dose of X chromosome gene expression in females and males.

Barr Bodies

Murray L. Barr and Ewart G. Bertram’s experiments with cats, as well as Keith Moore and Barr’s subsequent study in humans, demonstrate a genetic mechanism in mammals that compensates for X chromosome dosage disparities. Barr and Bertram observed a darkly staining body in the interphase nerve cells of female cats that was absent in similar cells of males. In humans, this body can be easily demonstrated in female cells derived from the buccal mucosa (cheek cells) or in fibroblasts (undifferentiated connective tissue cells), but not in similar male cells (Figure 5.4). This highly condensed structure, about 1 μm in diameter, lies against the nuclear envelope of interphase cells, and it stains positively for a number of different DNA-binding dyes.

This chromosome structure, called a **sex chromatin body**, or simply a **Barr body**, is an inactivated X chromosome. Susumu Ohno was the first to suggest that the Barr body arises from one of the two X chromosomes. This hypothesis is attractive because it provides a possible mechanism for dosage compensation. If one of the two X chromosomes is inactive in the cells of females, the dosage of genetic information that can be expressed in males and females will be equivalent. Convincing, though indirect, evidence for this hypothesis comes from study of the sex-chromosome syndromes described earlier in this chapter. Regardless of how many X chromosomes a somatic cell possesses, all but one of them appear to be inactivated and can be seen as Barr bodies. For example, no Barr body is seen in the somatic cells of Turner 45,X females; one is seen in Klinefelter 47,XXY males; two in 47,XXX females; three in 48,XXXX females; and so on (Figure 5.5). Therefore, the number of Barr bodies follows an $N - 1$ rule, where N is the total number of X chromosomes present.

Although this apparent inactivation of all but one X chromosome increases our understanding of dosage compensation, it further complicates our perception of other

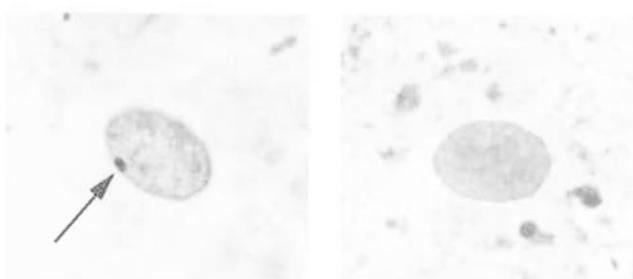


FIGURE 5.4 Photomicrographs comparing cheek epithelial cell nuclei from a male that fails to reveal Barr bodies (right) with a nucleus from a female that demonstrates a Barr body (indicated by the arrow in the left image). This structure, also called a sex chromatin body, represents an inactivated X chromosome.

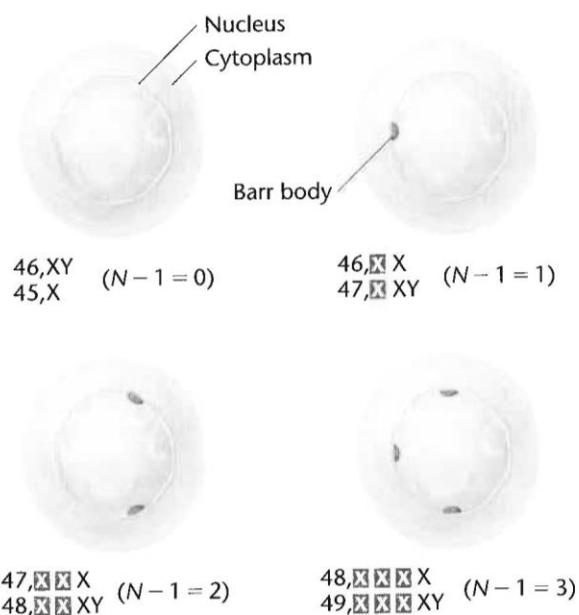


FIGURE 5.5 Occurrence of Barr bodies in various human karyotypes, where all X chromosomes except one ($N - 1$) are inactivated.

matters. For example, because one of the two X chromosomes is inactivated in normal human females, why then is the Turner 45,X individual not entirely normal? Why aren't females with the triplo-X and tetra-X karyotypes (47,XXX and 48,XXXX) completely unaffected by the additional X chromosome? Furthermore, in Klinefelter syndrome (47,XXY), X chromosome inactivation effectively renders the person 46,XY. Why aren't these males unaffected by the extra X chromosome in their nuclei?

One possible explanation is that chromosome inactivation does not normally occur in the very early stages of development of those cells destined to form gonadal tissues.

Another possible explanation is that not all genes on each X chromosome forming a Barr body are inactivated. Recent studies have indeed demonstrated that as many as 15 percent of the human X chromosomal genes actually escape inactivation. Clearly, then, not every gene on the X requires inactivation. In either case, excessive expression of certain X-linked genes might still occur at critical times during development despite apparent inactivation of superfluous X chromosomes.

The Lyon Hypothesis

In mammalian females, one X chromosome is of maternal origin, and the other is of paternal origin. Which one is inactivated? Is the inactivation random? Is the same chromosome inactive in all somatic cells? In the early 1960s, Mary Lyon, Liane Russell, and Ernest Beutler independently proposed a hypothesis that answers these questions. They postulated that the inactivation of X chromosomes occurs randomly in somatic cells at a point early in embryonic development, most likely sometime during the blastocyst stage of development. Once inactivation has occurred, all descendant cells have the same X chromosome inactivated as their initial progenitor cell.

This explanation, which has come to be called the **Lyon hypothesis**, was initially based on observations of female mice heterozygous for X-linked coat-color genes. The pigmentation of these heterozygous females was mottled, with large patches expressing the color allele on one X and other patches expressing the allele on the other X. This is the phenotypic pattern that would be expected if different X chromosomes were inactive in adjacent patches of cells. Similar mosaic patterns occur in the black and yellow-orange patches of female tortoiseshell and calico cats (Figure 5.6). Such X-linked coat-color patterns do not

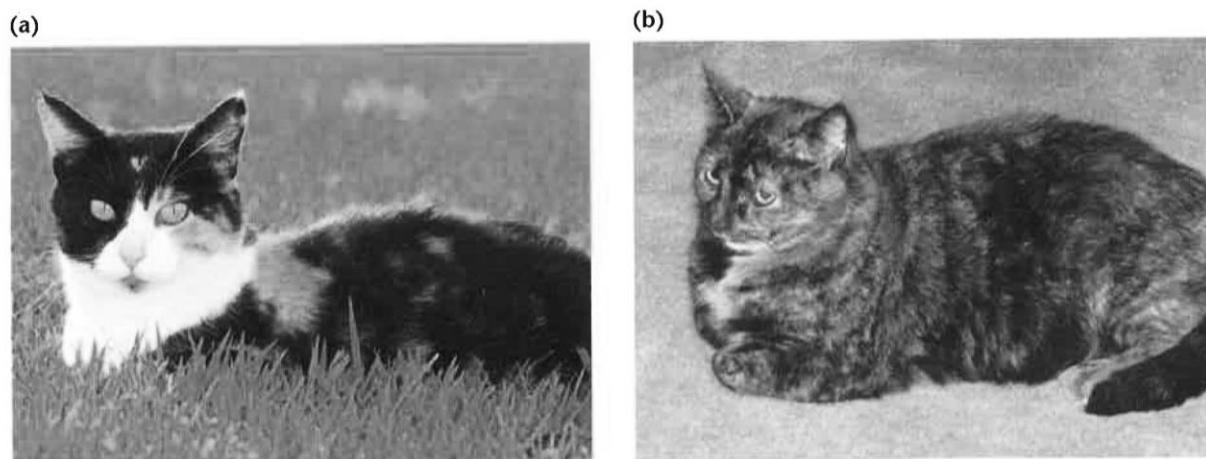


FIGURE 5.6 (a) The random distribution of orange and black patches in a calico cat illustrates the Lyon hypothesis. The white patches are due to another gene, distinguishing calico cats from tortoiseshell cats (b), which lack the white patches.

occur in male cats because all their cells contain the single maternal X chromosome and are therefore hemizygous for only one X-linked coat-color allele.

The most direct evidence in support of the Lyon hypothesis comes from studies of gene expression in clones of human fibroblast cells. Individual cells are isolated following biopsy and cultured *in vitro*. A culture of cells derived from a single cell is called a **clone**. The synthesis of the enzyme glucose-6-phosphate dehydrogenase (G6PD) is controlled by an X-linked gene. Numerous mutant alleles of this gene have been detected, and their gene products can be differentiated from the wild-type enzyme by their migration pattern in an electrophoretic field.

Fibroblasts have been taken from females heterozygous for different allelic forms of *G6PD* and studied. The Lyon hypothesis predicts that if inactivation of an X chromosome occurs randomly early in development, and thereafter all progeny cells have the same X chromosome inactivated as their progenitor, such a female should show two types of clones, each containing only one electrophoretic form of *G6PD*, in approximately equal proportions. This prediction has been confirmed experimentally, and studies involving modern techniques in molecular biology have clearly established that X chromosome inactivation occurs.

One ramification of X-inactivation is that mammalian females are mosaics for all heterozygous X-linked alleles—some areas of the body express only the maternally derived alleles, and others express only the paternally derived alleles. An especially interesting example involves **red-green color blindness**, an X-linked recessive disorder. In humans, hemizygous males are fully color-blind in all retinal cells. However, heterozygous females display mosaic retinas, with patches of defective color perception and surrounding areas with normal color perception. In this example, random inactivation of one or the other X chromosome early in the development of heterozygous females has led to these phenotypes.

The Mechanism of Inactivation

The least understood aspect of the Lyon hypothesis is the mechanism of X chromosome inactivation. Somehow, either DNA, the attached histone proteins, or both DNA and histone proteins, are chemically modified, silencing most genes that are part of that chromosome. Once silenced, a memory is created that keeps the same homolog inactivated following chromosome replications and cell divisions. Such a process, whereby expression of genes on one homolog, but not the other, is affected, is referred to as **imprinting**. This term also applies to a number of other examples in which genetic information is modified and gene expression is repressed. Collectively, such events are part of the growing field of **epigenetics** (see Chapter 19).

NOW SOLVE THIS

5.2 Carbon Copy (CC), the first cat produced from a clone, was created from an ovarian cell taken from her genetic donor, Rainbow, a calico cat. The diploid nucleus from the cell was extracted and then injected into an enucleated egg. The resulting zygote was then allowed to develop in a petri dish, and the cloned embryo was implanted in the uterus of a surrogate mother cat, who gave birth to CC. CC's surrogate mother was a tabby (see the photo below). Geneticists were very interested in the outcome of cloning a calico cat because they were not certain if the cloned cat would have patches of orange and black, just orange, or just black. Taking into account the Lyon hypothesis, explain the basis of the uncertainty. Would you expect CC to appear identical to Rainbow? Explain why or why not.



Carbon Copy with her surrogate mother.

■ **HINT:** This problem involves an understanding of the Lyon hypothesis. The key to its solution is to realize that the donor nucleus was from a differentiated ovarian cell of an adult female cat, which itself had inactivated one of its X chromosomes.

Ongoing investigations are beginning to clarify the mechanism of inactivation. A region of the mammalian X chromosome is the major control unit. This region, located on the proximal end of the p arm in humans, is called the **X-inactivation center (Xic)**, and its genetic expression occurs only on the X chromosome that is inactivated. The Xic is about 1 Mb (10^6 base pairs) in length and is known to contain several putative regulatory units and four genes. One of these, *X-inactive specific transcript (XIST)*, which encodes a long noncoding RNA (lncRNA), is now known to be a critical gene for X-inactivation.

Interesting observations have been made regarding the XIST lncRNA, many coming from experiments that focused on the equivalent gene in the mouse (*Xist*). First, the RNA product is quite large and does not encode a protein, and thus is not translated. The *Xist* transcript is an example of a lncRNA. An *Xist* lncRNA recruits a protein complex that

silences transcription at the epigenetic level. The role of lncRNAs in gene expression regulation will be discussed in greater detail in Chapter 19. The *Xist* lncRNAs spread over and coat the X chromosome bearing the gene that produced them. Two other noncoding genes at the *Xic* locus, *Tsix* (an antisense partner of *Xist*) and *Xite*, are also believed to play important roles in X-inactivation. It is thought that both *Xist* lncRNA expression and X-linked gene silencing are controlled by a small number of genes that are not inactivated in females.

Another observation is that transcription of *Xist* initially occurs at low levels on all X chromosomes. As the inactivation process begins, however, transcription continues, and is enhanced, only on the X chromosome that becomes inactivated. In addition to silencing genes on the inactivated X chromosome, there are 3D changes in the structure of the inactivated X chromosome that exclude RNA polymerase II from binding to transcription complexes. Recent work has revealed that the *Xist* lncRNA plays a role in these 3D changes in chromosome structure, which in turn help *Xist* to spread and inactivate genes across the X chromosome.

Interesting questions remain regarding imprinting and inactivation. For example, in cells with more than two X chromosomes, what sort of “counting” mechanism exists that designates all but one X chromosome to be inactivated? Studies by Jeannie T. Lee and colleagues suggest that maternal and paternal X chromosomes must first pair briefly and align at their *Xic* loci as a mechanism for counting the number of X chromosomes prior to X-inactivation [Figure 5.7(a)]. Using mouse embryonic stem cells, Lee’s group deleted the *Tsix* gene contained in the *Xic* locus. This

deletion blocked X–X pairing and resulted in chaotic inactivation of 0, 1, or 2 X chromosomes [Figure 5.7(b)]. Lee and colleagues provided further evidence for the role of the *Xic* locus in chromosome counting by adding copies of genetically engineered non-X chromosomes containing multiple copies of *Tsix* or *Xite*. (These are referred to as **transgenes** because they are artificially introduced into the organism.) This experimental procedure effectively blocked X–X pairing and prevented X chromosome inactivation [Figure 5.7(c)].

Other genes and protein products are being examined for their role in X chromosome pairing and counting. Recent studies by Lee and colleagues have provided evidence that the inactivated X chromosome must associate with regions at the periphery of the nucleus to maintain a state of silenced gene expression. Indeed, in a majority of human female somatic cells the inactivated X chromosome, present as a Barr body, is observed attached to the nuclear envelope.

Many questions remain. What “blocks” the *Xic* locus of the active chromosome, preventing further transcription of *Xist*? How does imprinting impart a memory such that inactivation of the same X chromosome or chromosomes is subsequently maintained in progeny cells, as the Lyon hypothesis calls for? Whatever the answers to these questions, scientists have taken exciting steps toward understanding how dosage compensation is accomplished in mammals.

Finally, modern applications of genomic analysis (which you will learn more about in Chapter 22) have enabled researchers to compare gene expression changes between males and females on a genome-wide scale. Such

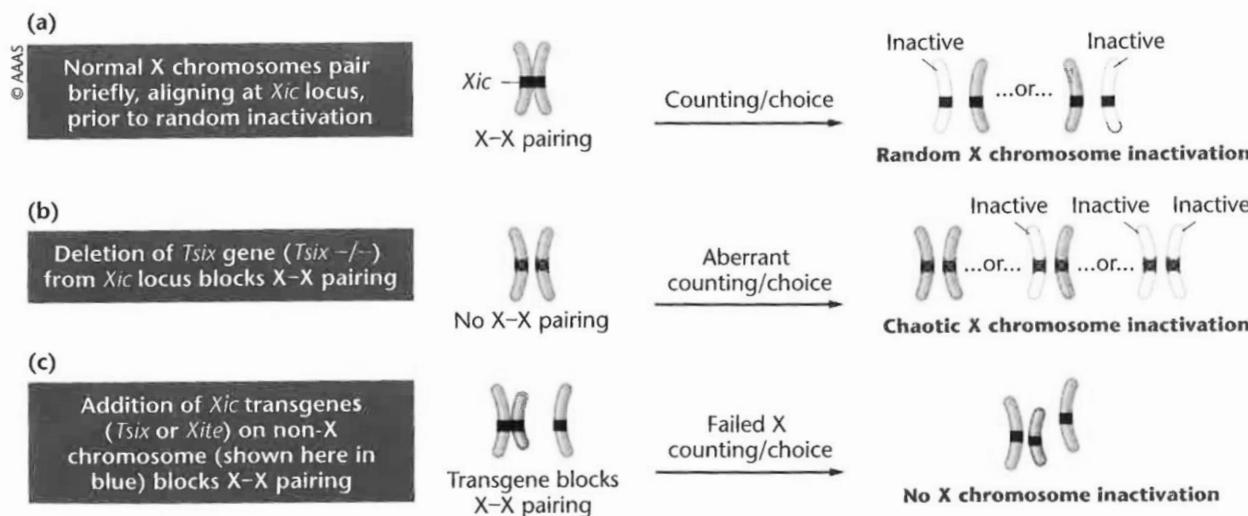


FIGURE 5.7 (a) Transient pairing of X chromosomes may be required for initiating X-inactivation. (b) Deleting the *Tsix* gene of the *Xic* locus prevents X–X pairing and leads to chaotic X-inactivation. (c) Blocking X–X pairing by addition of *Xic*-containing transgenes blocks X–X pairing and prevents X-inactivation.

work is revealing examples of sex-biased gene expression—genes that are expressed predominantly in one sex or another (or at a higher level in one sex). Undoubtedly these studies will provide additional insight about silenced genes and activated genes that contribute to sex determination.

5.5 The Ratio of X Chromosomes to Sets of Autosomes Can Determine Sex

We now discuss two interesting cases where the Y chromosome does not play a role in sex determination. First, in the fruit fly, *Drosophila melanogaster*, even though most males contain a Y chromosome, the Y plays no role. Second, in the roundworm, *Caenorhabditis elegans*, the organism lacks a Y chromosome altogether. In both cases, we shall see that the critical factor is the ratio of X chromosomes to the number of sets of autosomes.

D. melanogaster

Because males and females in *Drosophila melanogaster* (and other *Drosophila* species) have the same general sex-chromosome composition as humans (males are XY and females are XX), we might assume that the Y chromosome also causes maleness in these flies. However, the elegant work of Calvin Bridges in 1921 showed this not to be true. His studies of flies with quite varied chromosome compositions led him to the conclusion that the Y chromosome is not involved in sex determination in this organism. Instead, Bridges proposed that the X chromosomes and autosomes together play a critical role in sex determination.

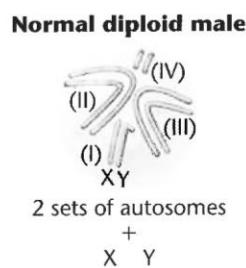
Bridges' work can be divided into two phases: (1) A study of offspring resulting from nondisjunction of the X chromosomes during meiosis in females and (2) subsequent work with progeny of females containing three copies of each chromosome, called triploid ($3n$) females. As we have seen previously in this chapter (and as you will see in Figure 6.1), nondisjunction is the failure of paired chromosomes to segregate or separate during the anaphase stage of the first or second meiotic divisions. The result is the production of two types of abnormal gametes, one of which contains an extra chromosome ($n + 1$) and the other of which lacks a chromosome ($n - 1$). Fertilization of such gametes with a haploid gamete produces $(2n + 1)$ or $(2n - 1)$ zygotes. As in humans, if nondisjunction involves the X chromosome, in addition

to the normal complement of autosomes, both an XXY and an XO sex-chromosome composition may result. (The "0" signifies that neither a second X nor a Y chromosome is present, as occurs in XO genotypes of individuals with Turner syndrome.)

Contrary to what was later discovered in humans, Bridges found that the XXY flies were normal females and the XO flies were sterile males. The presence of the Y chromosome in the XXY flies did not cause maleness, and its absence in the XO flies did not produce femaleness. From these data, Bridges concluded that the Y chromosome in *Drosophila* lacks male-determining factors, but since the XO males were sterile, it does contain genetic information essential to male fertility.

Bridges was able to clarify the mode of sex determination in *Drosophila* by studying the progeny of triploid females ($3n$), which have three copies each of the haploid complement of chromosomes. *Drosophila* has a haploid number of 4, thereby possessing three pairs of autosomes in addition to its pair of sex chromosomes. Triploid females apparently originate from rare diploid eggs fertilized by normal haploid sperm. Triploid females have heavy-set bodies, coarse bristles, and coarse eyes, and they may be fertile. Because of the odd number of each chromosome (3), during meiosis, a variety of different chromosome complements are distributed into gametes that give rise to offspring with a variety of abnormal chromosome constitutions. Correlations between the sexual morphology and chromosome composition, along with Bridges' interpretation, are shown in Figure 5.8.

Bridges realized that the critical factor in determining sex is the ratio of X chromosomes to the number of haploid



Chromosome formulation	Ratio of X chromosomes to autosome sets	Sexual morphology
3X:2A	1.5	Metafemale
3X:3A	1.0	Female
2X:2A	1.0	Female
3X:4A	0.75	Intersex
2X:3A	0.67	Intersex
X:2A	0.50	Male
XY:2A	0.50	Male
XY:3A	0.33	Metamale

FIGURE 5.8 The ratios of X chromosomes to sets of autosomes and the resultant sexual morphology seen in *Drosophila melanogaster*.

sets of autosomes (A) present. Normal (2X:2A) and triploid (3X:3A) females each have a ratio equal to 1.0, and both are fertile. As the ratio exceeds unity (3X:2A, or 1.5, for example), what was once called a *superfemale* is produced. Because such females are most often inviable, they are now more appropriately called **metafemales**.

Normal (XY:2A) and sterile (X0:2A) males each have a ratio of 1:2, or 0.5. When the ratio decreases to 1:3, or 0.33, as in the case of an XY:3A male, infertile **metamales** result. Other flies recovered by Bridges in these studies had an (X:A) ratio intermediate between 0.5 and 1.0. These flies were generally larger, and they exhibited a variety of morphological abnormalities and rudimentary bisexual gonads and genitalia. They were invariably sterile and expressed both male and female morphology, thus being designated as **intersexes**.

Bridges' results indicate that in *Drosophila*, factors that cause a fly to develop into a male are not located on the sex chromosomes but are instead found on the autosomes. Some female-determining factors, however, are located on the X chromosomes. Thus, with respect to primary sex determination, male gametes containing one of each autosome plus a Y chromosome result in male offspring not because of the presence of the Y but because they fail to contribute an X chromosome. This mode of sex determination is explained by the **genic balance theory**. Bridges proposed that a threshold for maleness is reached when the X:A ratio is 1:2 (X:2A), but that the presence of an additional X (XX:2A) alters the balance and results in female differentiation.

Numerous genes involved in sex determination in *Drosophila* have been identified. The recessive autosomal gene *transformer* (*tra*), discovered over 50 years ago by Alfred H. Sturtevant, clearly demonstrated that a single autosomal gene could have a profound impact on sex determination. Females homozygous for *tra* are transformed into sterile males, but homozygous males are unaffected. More recently, another gene, *Sex-lethal* (*Sxl*), has been shown to play a critical role, serving as a "master switch" in sex determination. Activation of the X-linked *Sxl* gene, which relies on a ratio of X chromosomes to sets of autosomes that equals 1.0, is essential to female development. In the absence of activation—as when, for example, the X:A ratio is 0.5—male development occurs.

Although it is not yet exactly clear how this ratio influences the *Sxl* locus, we do have some insights into the question. The *Sxl* locus is part of a hierarchy of gene expression and exerts control over other genes, including *tra* (discussed in the previous paragraph) and *dsx* (*doublesex*). The wild-type allele of *tra* is activated by the product of *Sxl* only in females and in turn influences the expression of *dsx*. Depending on how the initial RNA transcript of *dsx* is processed the resultant *dsx* protein activates either male- or female-specific genes required for sexual differentiation.

It is interesting to note that *dsx* homologs have been found in all mammals, thus emphasizing a role for this gene in gonad development.

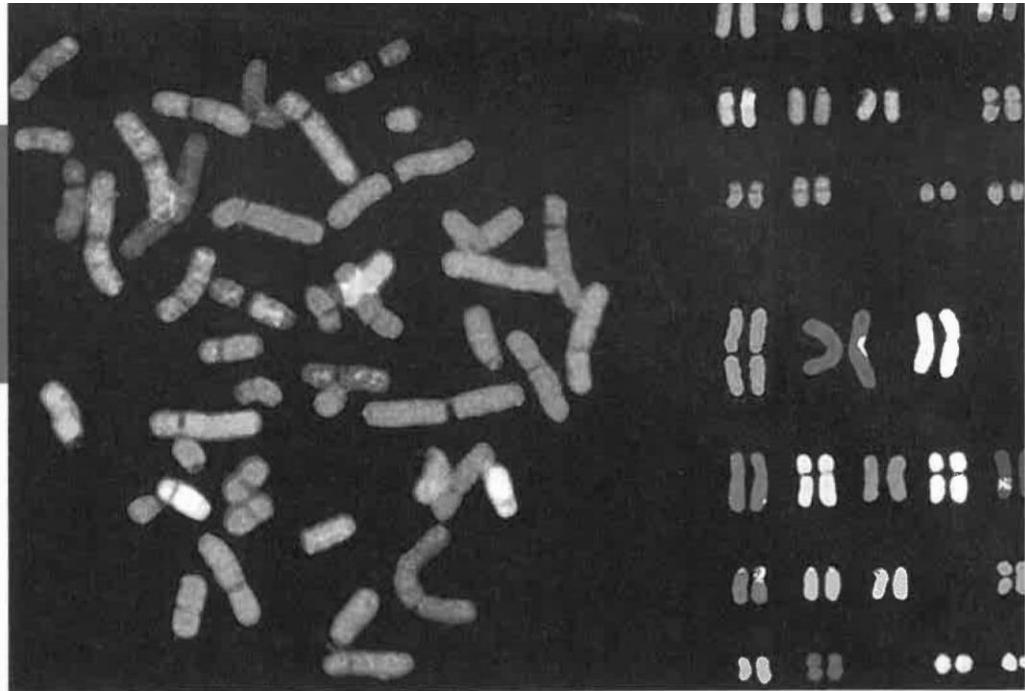
Each step in this regulatory cascade requires a form of processing called **RNA splicing**, in which portions of the RNA are removed and the remaining fragments are "spliced" back together prior to translation into a protein. In the case of the *Sxl* gene, the RNA transcript may be spliced in different ways, a phenomenon called **alternative splicing**. A different RNA transcript is produced in females than in males. In potential females, the transcript is active and initiates a cascade of regulatory gene expression, ultimately leading to female differentiation. In potential males, the transcript is inactive, leading to a different pattern of gene activity, whereby male differentiation occurs. We will return to this topic in Chapter 18, where alternative splicing is again addressed as one of the mechanisms involved in the regulation of genetic expression in eukaryotes.

1.2.6

Klug et al. (2019)

Page 151-154, and 156-161

6



Chromosomal Mutations: Variation in Number and Arrangement

CHAPTER CONCEPTS

- The failure of chromosomes to properly separate during meiosis results in variation in the chromosome content of gametes and subsequently in offspring arising from such gametes.
- Plants often tolerate an abnormal genetic content, but, as a result, they often manifest unique phenotypes. Such genetic variation has been an important factor in the evolution of plants.
- In animals, genetic information is in a delicate equilibrium whereby the gain or loss of a chromosome, or part of a chromosome, in an otherwise diploid organism often leads to lethality or to an abnormal phenotype.
- The rearrangement of genetic information within the genome of a diploid organism may be tolerated by that organism but may affect the viability of gametes and the phenotypes of organisms arising from those gametes.
- Chromosomes in humans contain fragile sites—regions susceptible to breakage, which lead to abnormal phenotypes.

In previous chapters, we have emphasized how mutations and the resulting alleles affect an organism's phenotype and how traits are passed from parents to offspring according to Mendelian principles. In this chapter, we look at phenotypic variation that results from

Spectral karyotyping of human chromosomes, utilizing differentially labeled "painting" probes.

more substantial changes than alterations of individual genes—modifications at the level of the chromosome.

Although most members of diploid species normally contain precisely two haploid chromosome sets, many known cases vary from this pattern. Modifications include a change in the total number of chromosomes, the deletion or duplication of genes or segments of a chromosome, and rearrangements of the genetic material either within or among chromosomes. Taken together, such changes are called **chromosome mutations** or **chromosome aberrations**, to distinguish them from gene mutations. Because the chromosome is the unit of genetic transmission, according to Mendelian laws, chromosome aberrations are passed to offspring in a predictable manner, resulting in many unique genetic outcomes.

Because the genetic component of an organism is delicately balanced, even minor alterations of either content or location of genetic information within the genome can result in some form of phenotypic variation. More substantial changes may be lethal, particularly in animals. Throughout this chapter, we consider many types of chromosomal aberrations, the phenotypic consequences for the organism that harbors an aberration, and the impact of the aberration on the offspring of an affected individual. We will also discuss the role of chromosomal aberrations in the evolutionary process.

6.1 Variation in Chromosome Number: Terminology and Origin

Variation in chromosome number ranges from the addition or loss of one or more chromosomes to the addition of one or more haploid sets of chromosomes. Before we embark on our discussion, it is useful to clarify the terminology that describes such changes. In the general condition known as **aneuploidy**, an organism gains or loses one or more chromosomes but not a complete set. The loss of a single chromosome from an otherwise diploid genome is called *monosomy*. The gain of one chromosome results in *trisomy*. These changes are contrasted with the condition of **euploidy**, where complete haploid sets of chromosomes are present. If more than two sets are present, the term **polyploidy** applies. Organisms with three sets are specifically *triploid*, those with four sets are *tetraploid*, and so on. **Table 6.1** provides an organizational framework for you to follow as we discuss each of these categories of aneuploid and euploid variation and the subsets within them.

As we consider cases that include the gain or loss of chromosomes, it is useful to examine how such aberrations originate. For instance, how do the syndromes arise where the number of sex-determining chromosomes in humans is altered (Chapter 5)? As you may recall, the gain (47,XXY) or loss (45,X) of an X chromosome from an otherwise diploid genome affects the phenotype, resulting in **Klinefelter syndrome** or **Turner syndrome**, respectively (see Figure 5.2). Human females may contain extra X chromosomes (e.g., 47,XXX, 48,XXXX), and some males contain an extra Y chromosome (47,XY).

Such chromosomal variation originates as a random error during the production of gametes, a phenomenon referred to as **nondisjunction**, whereby paired homologs

TABLE 6.1 Terminology for Variation in Chromosome Numbers

Term	Explanation
Aneuploidy	$2n \pm x$ chromosomes
Monosomy	$2n - 1$
Disomy	$2n$
Trisomy	$2n + 1$
Tetrasomy, pentasomy, etc.	$2n + 2, 2n + 3$, etc.
Euploidy	Multiples of n
Diploidy	$2n$
Polyploidy	$3n, 4n, 5n, \dots$
Tripliody	$3n$
Tetraploidy, pentaploidy, etc.	$4n, 5n$, etc.
Autopolyploidy	Multiples of the same genome
Allopolyploidy (amphidiploidy)	Multiples of closely related genomes

NOW SOLVE THIS

6.1 A human female with Turner syndrome (45,X) also expresses the X-linked trait hemophilia, as did her father. Which of her parents underwent nondisjunction during meiosis, giving rise to the gamete responsible for the syndrome?

■ **HINT:** This problem involves an understanding of how nondisjunction leads to aneuploidy. The key to its solution is first to review Turner syndrome, discussed above and in more detail later in the text (see Chapter 5), then to factor in that she expresses hemophilia, and finally, to consider which parent contributed a gamete with an X chromosome that underwent normal meiosis.

For more practice, see Problems 27 and 28.

fail to disjoin during segregation. This process disrupts the normal distribution of chromosomes into gametes. The results of nondisjunction during meiosis I and meiosis II for a single chromosome of a diploid organism are shown in **Figure 6.1**. As you can see, abnormal gametes can form containing either two members of the affected chromosome or none at all. Fertilizing these with a normal haploid gamete produces a zygote with either three members (trisomy) or only one member (monosomy) of this chromosome. Nondisjunction leads to a variety of aneuploid conditions in humans and other organisms.

6.2 Monosomy and Trisomy Result in a Variety of Phenotypic Effects

We turn now to a consideration of variations in the number of autosomes and the genetic consequence of such changes. The most common examples of **aneuploidy**, where an organism has a chromosome number other than an exact multiple of the haploid set, are cases in which a single chromosome is either added to, or lost from, a normal diploid set.

Monosomy

The loss of one chromosome produces a $2n - 1$ complement called **monosomy**. Although monosomy for the X chromosome occurs in humans, as we have seen in 45,X Turner syndrome, monosomy for any of the autosomes is not usually tolerated in humans or other animals. In *Drosophila*, flies that are monosomic for the very small chromosome IV (containing less than 5 percent of the organism's genes) develop more slowly, exhibit reduced body size, and have impaired viability. Monosomy for the larger autosomal chromosomes II and III is apparently lethal because such flies have never been recovered.

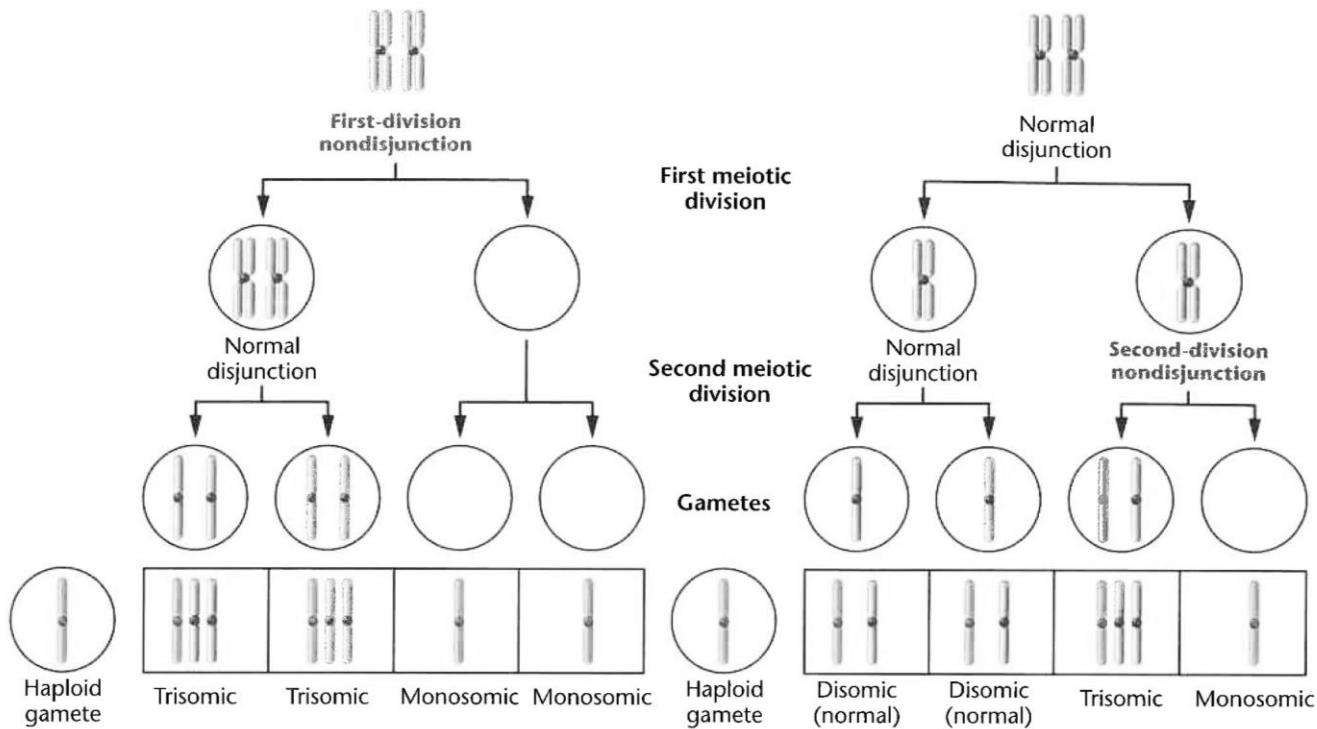


FIGURE 6.1 Nondisjunction during the first and second meiotic divisions. In both cases, some of the gametes that are formed either contain two members of a specific chromosome or lack that chromosome. After fertilization by a gamete with a normal haploid content, monosomic, disomic (normal), or trisomic zygotes are produced.

The failure of monosomic individuals to survive is at first quite puzzling, since at least a single copy of every gene is present in the remaining homolog. However, if just one of those genes is represented by a lethal allele, the unpaired chromosome condition will result in the death of the organism. This will occur because monosity unmasks recessive lethals that are otherwise tolerated in heterozygotes carrying the corresponding wild-type alleles. Another possible cause of lethality of aneuploidy is that a single copy of a recessive gene may be insufficient to provide adequate function for sustaining the organism, a phenomenon called **haploinsufficiency**.

Aneuploidy is better tolerated in the plant kingdom. Monosity for autosomal chromosomes has been observed in maize, tobacco, the evening primrose (*Oenothera*), and the jimson weed (*Datura*), among many other plants. Nevertheless, such monosomic plants are usually less viable than their diploid derivatives. Haploid pollen grains, which undergo extensive development before participating in fertilization, are particularly sensitive to the lack of one chromosome and are seldom viable.

Trisomy

In general, the effects of **trisomy** ($2n + 1$) parallel those of monosity. However, the addition of an extra chromosome produces somewhat more viable individuals in both animal

and plant species than does the loss of a chromosome. In animals, this is often true, provided that the chromosome involved is relatively small. However, the addition of a large autosome to the diploid complement in both *Drosophila* and humans has severe effects and is usually lethal during development.

In plants, trisomic individuals are viable, but their phenotype may be altered. A classical example involves the jimson weed, *Datura*, whose diploid number is 24. Twelve primary trisomic conditions are possible, and examples of each one have been recovered. Each trisomy alters the phenotype of the plant's capsule (Figure 6.2) sufficiently to produce a unique phenotype. These capsule phenotypes were first thought to be caused by mutations in one or more genes.

Still another example is seen in the rice plant (*Oryza sativa*), which has a haploid number of 12. Trisomic strains for each chromosome have been isolated and studied—the plants of 11 strains can be distinguished from one another and from wild-type plants. Trisomics for the longer chromosomes are the most distinctive, and the plants grow more slowly. This is in keeping with the belief that larger chromosomes cause greater genetic imbalance than smaller ones. Leaf structure, foliage, stems, grain morphology, and plant height also vary among the various trisomies.



FIGURE 6.2 The capsule of the fruit of the jimson weed, *Datura stramonium*, the phenotype of which is uniquely altered by each of the possible 12 trisomic conditions.

Down Syndrome: Trisomy 21

The only human autosomal trisomy in which a significant number of individuals survive longer than a year past birth was discovered in 1866 by Langdon Down. The condition is now known to result from trisomy of chromosome 21 (Figure 6.3) and is called **Down syndrome** or simply **trisomy 21** (designated 47,21+). This trisomy is found in approximately 1 infant in every 800 live births. While this might seem to be a rare, improbable event, there are approximately 4000–5000 such births annually in the United States, and there are currently over 250,000 individuals with Down syndrome.

Typical of other conditions classified as syndromes, many phenotypic characteristics *may* be present in trisomy

21, but any single affected individual usually exhibits only a subset of these. In the case of Down syndrome, there are 12 to 14 such characteristics, with each individual, on average, expressing 6 to 8 of them. Nevertheless, the outward appearance of these individuals is very similar, and they bear a striking resemblance to one another. This is largely due to a prominent epicanthic fold in each eye* and the typically flat face and round head. Those with Down syndrome are also characteristically short and may have a protruding, furrowed tongue (which causes the mouth to remain partially open) and short, broad hands with characteristic palm and finger-print patterns. Physical, psychomotor, and mental development is retarded, and poor muscle tone is characteristic. While life expectancy is shortened to an average of about 50 years, individuals are known to survive into their 60s.

Children afflicted with Down syndrome are also prone to respiratory disease and heart malformations, and they show an incidence of leukemia approximately 20 times higher than that of the normal population. In addition, death in older Down syndrome adults is frequently due to Alzheimer disease, the onset of which occurs at a much earlier age than in the normal population.

The Down Syndrome Critical Region (DSCR)

Because Down syndrome is common in our population, a comprehensive understanding of the underlying genetic basis has long been a research goal. Investigations have

*The epicanthic fold, or epicanthus, is a skin fold of the upper eyelid, extending from the nose to the inner side of the eyebrow. It covers and appears to lower the inner corner of the eye, giving the eye a slanted, or almond-shaped, appearance. The epicanthus is a prominent normal component of the eyes in many Asian groups.



FIGURE 6.3 The karyotype and a photograph of a child with Down syndrome (hugging her unaffected sister on the right). In the karyotype, three members of chromosome 21 are present, creating the 47,21+ condition.

given rise to the idea that a critical region of chromosome 21 contains the genes that are dosage sensitive in this trisomy and responsible for the many phenotypes associated with the syndrome. This hypothetical portion of the chromosome has been called the **Down syndrome critical region (DSCR)**. A mouse model was created in 2004 that is trisomic for the DSCR, although some mice do not exhibit the characteristics of the syndrome. Nevertheless, this remains an important investigative approach, which is explored in the Modern Approaches to Understanding Gene Function box.

Current studies of the DSCR in both humans and mice have led to several interesting findings. Research investigations have now led us to believe that the presence of three copies of the genes present in this region is necessary, but not sufficient for the cognitive deficiencies characteristic of the syndrome. Another finding involves an important observation about Down syndrome—that such individuals have a decreased risk of developing a number of cancers involving solid tumors, including lung cancer and melanoma. This health benefit has been correlated with the presence of an extra copy of the *DSCR1* gene, which encodes a protein that suppresses *vascular endothelial growth factor (VEGF)*. This suppression, in turn, blocks the process of angiogenesis. As a result, the overexpression of this gene inhibits tumors from forming proper vascularization, diminishing their growth. A 14-year study published in 2002 involving 17,800 individuals with Down syndrome revealed an approximate 10 percent reduction in cancer mortality in contrast to a control population. No doubt, further information will be forthcoming from the study of the DSCR.

The Origin of the Extra Chromosome 21 in Down Syndrome

Most frequently, this trisomic condition occurs through nondisjunction of chromosome 21 during meiosis. Failure of paired homologs to disjoin during either anaphase I or II may lead to gametes with the $n + 1$ chromosome composition. About 75 percent of these errors leading to Down syndrome are attributed to nondisjunction during meiosis I. Subsequent fertilization with a normal gamete creates the trisomic condition.

Chromosome analysis has shown that, while the additional chromosome may be derived from either the mother or father, the ovum is the source in about 95 percent of 47,21+ trisomy cases. Before the development of techniques using polymorphic markers to distinguish paternal from maternal homologs, this conclusion was supported by the more indirect evidence derived from studies of the age of mothers giving birth to infants afflicted with Down

syndrome. **Figure 6.4** shows the relationship between the incidence of Down syndrome births and maternal age, illustrating the dramatic increase as the age of the mother increases.

While the frequency is about 1 in 1000 at maternal age 30, a tenfold increase to a frequency of 1 in 100 is noted at age 40. The frequency increases still further to about 1 in 30 at age 45. A very alarming statistic is that as the age of childbearing women exceeds 45, the probability of a Down syndrome birth continues to increase substantially. In spite of these statistics, substantially more than half of Down syndrome births occur to women younger than 35 years, because the overwhelming proportion of pregnancies in the general population involve women under that age.

Although the nondisjunctional event clearly increases with age, we do not know with certainty why this is so. However, one observation is clearly relevant. Meiosis is initiated in all the eggs of a human female when she is still a fetus, until the point where the homologs synapse and recombination has begun. Then oocyte development is arrested in meiosis I. Thus, all primary oocytes have been formed by birth. When ovulation begins at puberty, meiosis is reinitiated in one egg during each ovulatory cycle and continues into meiosis II. The process is once again arrested after ovulation and is not completed unless fertilization occurs.

The end result of this progression is that each ovum that is released has been arrested in meiosis I for about a month longer than the one released during the preceding cycle. As a consequence, women 30 or 40 years old produce ova that are significantly older and that have been

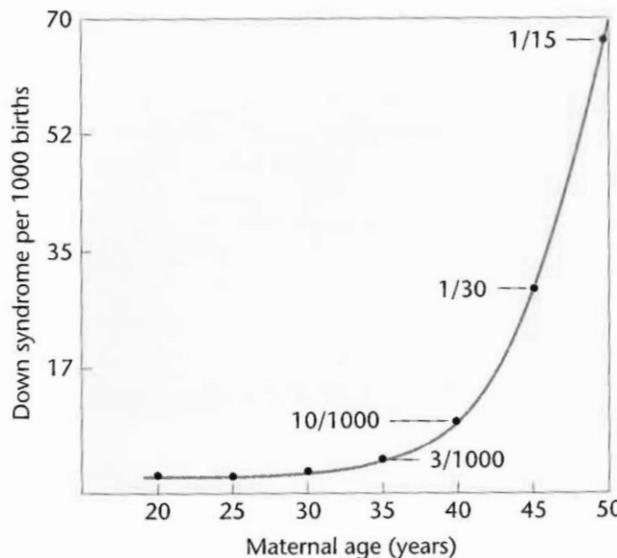


FIGURE 6.4 Incidence of Down syndrome births related to maternal age.

arrested longer than those they ovulated 10 or 20 years previously. In spite of the logic underlying this hypothesis explaining the cause of the increased incidence of Down syndrome as women age, it remains difficult to prove directly.

These statistics obviously pose a serious problem for the woman who becomes pregnant late in her reproductive years. Genetic counseling early in such pregnancies is highly recommended. Counseling informs prospective parents about the probability that their child will be affected and educates them about Down syndrome. Although some individuals with Down syndrome must be institutionalized, others benefit greatly from special education programs and may be cared for at home. (Down syndrome children in general are noted for their affectionate, loving nature.) A genetic counselor may also recommend a prenatal diagnostic technique in which fetal cells are isolated and cultured.

In **amniocentesis** and **chorionic villus sampling (CVS)**, the two most familiar approaches, fetal cells are obtained from the amniotic fluid or the chorion of the placenta, respectively. In a newer approach, fetal cells and DNA are derived directly from the maternal circulation, a technique referred to as **noninvasive prenatal genetic diagnosis (NIPGD)**. Requiring only a 10-mL maternal blood sample, this procedure will become increasingly more common because it poses no risk to the fetus. After fetal cells are obtained and cultured, the karyotype can be determined by cytogenetic analysis. If the fetus is diagnosed as being affected, a therapeutic abortion is one option currently available to parents. Obviously, this is a difficult decision involving a number of religious and ethical issues.

Since Down syndrome is caused by a random error—nondisjunction of chromosome 21 during maternal or paternal meiosis—the occurrence of the disorder is *not* expected to be inherited. Nevertheless, Down syndrome occasionally runs in families. These instances, referred to as *familial Down syndrome*, involve a translocation of chromosome 21, another type of chromosomal aberration, which we will discuss later in the chapter.

Human Aneuploidy

Besides Down syndrome, only two human trisomies, and no monosomies, survive to term: **Patau** and **Edwards syndromes** (47,13+ and 47,18+, respectively). Even so, these individuals manifest severe malformations and early lethality. Figure 6.5 illustrates the abnormal karyotype and the many defects characterizing Patau infants.

The above observation leads us to ask whether many other aneuploid conditions arise but that the affected

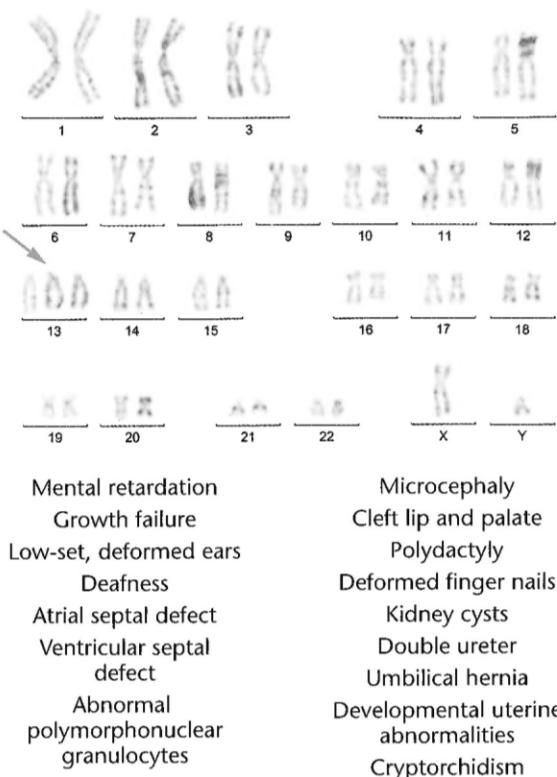


FIGURE 6.5 The karyotype and phenotypic description of an infant with Patau syndrome, where three members of chromosome 13 are present, creating the 47,13+ condition.

fetuses do not survive to term. That this is the case has been confirmed by karyotypic analysis of spontaneously aborted fetuses. These studies reveal two striking statistics: (1) Approximately 20 percent of all conceptions terminate in spontaneous abortion (some estimates are considerably higher); and (2) about 30 percent of all spontaneously aborted fetuses demonstrate some form of chromosomal imbalance. This suggests that at least 6 percent (0.20×0.30) of conceptions contain an abnormal chromosome complement. A large percentage of fetuses demonstrating chromosomal abnormalities are aneuploids.

An extensive review of this subject by David H. Carr has revealed that a significant percentage of aborted fetuses are trisomic for one of the chromosome groups. Trisomies for every human chromosome have been recovered. Interestingly, the monosomy with the highest incidence among abortuses is the 45,X condition, which produces an infant with Turner syndrome, if the fetus survives to term. Autosomal monosomies are seldom found, however, even though nondisjunction should produce $n - 1$ gametes with a frequency equal to $n + 1$ gametes. This finding suggests

that gametes lacking a single chromosome are functionally impaired to a serious degree or that the embryo dies so early in its development that recovery occurs infrequently. We discussed the potential causes of monosomic lethality earlier in this chapter. Carr's study also found various forms of polyploidy and other miscellaneous chromosomal anomalies.

These observations support the hypothesis that normal embryonic development requires a precise diploid complement of chromosomes to maintain the delicate equilibrium in the expression of genetic information. The prenatal mortality of most aneuploids provides a barrier against the introduction of these genetic anomalies into the human population.

6.3 Polyploidy, in Which More Than Two Haploid Sets of Chromosomes Are Present, Is Prevalent in Plants

The term *polyploidy* describes instances in which more than two multiples of the haploid chromosome set are found. The naming of polyploids is based on the number of sets of chromosomes found: A triploid has $3n$ chromosomes; a tetraploid has $4n$; a pentaploid, $5n$; and so forth (Table 6.1). Several general statements can be made about polyploidy. This condition is relatively infrequent in many animal species but is well known in lizards, amphibians, and fish, and is much more common in plant species. Usually, odd numbers of chromosome sets are not reliably maintained from generation to generation because a polyploid organism with an uneven number of homologs often does not produce genetically balanced gametes. For this reason,

triploids, pentaploids, and so on, are not usually found in plant species that depend solely on sexual reproduction for propagation.

Polyplody originates in two ways: (1) the addition of one or more extra sets of chromosomes, identical to the normal haploid complement of the same species, resulting in **autopolyploidy**; or (2) the combination of chromosome sets from different species occurring as a consequence of hybridization, resulting in **allopolyploidy** (from the Greek word *allo*, meaning "other" or "different"). The distinction between auto- and allopolyploidy is based on the genetic origin of the extra chromosome sets, as shown in Figure 6.6.

In our discussion of polyploidy, we use the following symbols to clarify the origin of additional chromosome sets. For example, if A represents the haploid set of chromosomes of any organism, then

$$A = a_1 + a_2 + a_3 + a_4 + \cdots + a_n$$

where a_1 , a_2 , and so on, are individual chromosomes and n is the haploid number. A normal diploid organism is represented simply as AA .

Autopolyploidy

In autopolyploidy, each additional set of chromosomes is identical to the parent species. Therefore, triploids are represented as AAA , tetraploids are $AAAA$, and so forth.

Autotriploids arise in several ways. A failure of all chromosomes to segregate during meiotic divisions can produce a diploid gamete. If such a gamete is fertilized by a haploid gamete, a zygote with three sets of chromosomes is produced. Or, rarely, two sperm may fertilize an ovum, resulting in a triploid zygote. Triploids are also produced under experimental conditions by crossing diploids with tetraploids. Diploid organisms produce

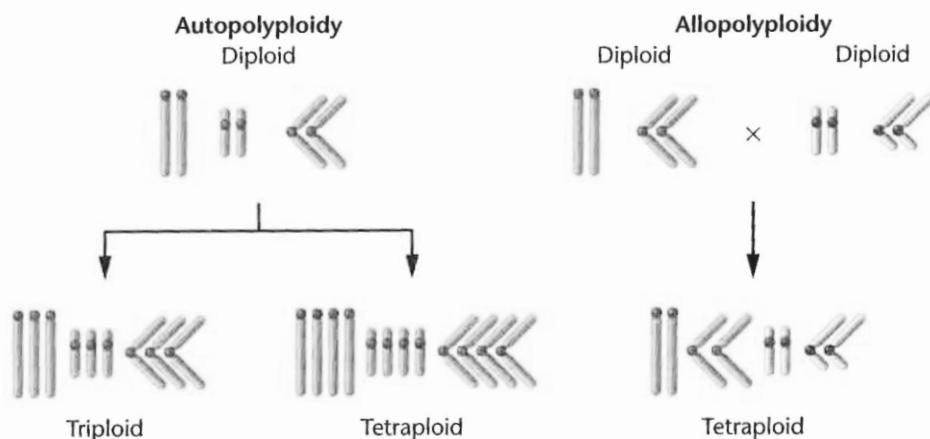


FIGURE 6.6 Contrasting chromosome origins of an autopolyploid versus an allopolyploid karyotype.

gametes with n chromosomes, while tetraploids produce $2n$ gametes. Upon fertilization, the desired triploid is produced.

Because they have an even number of chromosomes, **autotetraploids** ($4n$) are theoretically more likely to be found in nature than are autotriploids. Unlike triploids, which often produce genetically unbalanced gametes with odd numbers of chromosomes, tetraploids are more likely to produce balanced gametes when involved in sexual reproduction.

How polyploidy arises naturally is of great interest to geneticists. In theory, if chromosomes have replicated, but the parent cell never divides and instead reenters interphase, the chromosome number will be doubled. That this very likely occurs is supported by the observation that tetraploid cells can be produced experimentally from diploid cells. This process is accomplished by applying cold or heat shock to meiotic cells or by applying colchicine to somatic cells undergoing mitosis. **Colchicine**, an alkaloid derived from the autumn crocus, interferes with spindle formation, and thus replicated chromosomes cannot separate at anaphase and do not migrate to the poles. When colchicine is removed, the cell can reenter interphase. When the paired sister chromatids separate and uncoil, the nucleus contains twice the diploid number of chromosomes and is therefore $4n$. This process is shown in **Figure 6.7**.

In general, autopolyploids are larger than their diploid relatives. This increase seems to be due to larger cell size rather than greater cell number. Although autopolyploids do not contain new or unique information compared with their diploid relatives, the flower and fruit of plants are often increased in size, making such varieties of greater horticultural or commercial value. Economically important

triploid plants include several potato species of the genus *Solanum*, Winesap apples, commercial bananas, seedless watermelons, and the cultivated tiger lily *Lilium tigrinum*. These plants are propagated asexually. Diploid bananas contain hard seeds, but the commercial triploid “seedless” variety has edible seeds. Tetraploid alfalfa, coffee, peanuts, and McIntosh apples are also of economic value because they are either larger or grow more vigorously than do their diploid or triploid counterparts. Many of the most popular varieties of hosta plant are tetraploid. In each case, leaves are thicker and larger, the foliage is more vivid, and the plant grows more vigorously. The commercial strawberry is an octoploid.

We have long been curious about how cells with increased ploidy values, where no new genes are present, express different phenotypes from their diploid counterparts. Our current ability to examine gene expression using modern biotechnology has provided some interesting insights. For example, Gerald Fink and his colleagues have been able to create strains of the yeast *Saccharomyces cerevisiae* with one, two, three, or four copies of the genome. Thus, each strain contains identical genes (they are said to be isogenic) but different ploidy values. These scientists then examined the expression levels of all genes during the entire cell cycle of the organism. Using the rather stringent standards of a tenfold increase or decrease of gene expression, Fink and coworkers proceeded to identify ten cases where, as ploidy increased, gene expression was increased at least tenfold and seven cases where it was reduced by a similar level.

Among these cases are two genes that encode **G1 cyclins**, which are repressed as ploidy increases. These proteins facilitate the cell's movement through the cell cycle, which is thus delayed. The polyploid cell stays in the G1 stage longer than normal, and the cell grows to a larger size. Yeast cells also show different morphology as ploidy increases.

Several of the other genes, repressed as ploidy increases, have been linked to cytoskeletal dynamics that account for the morphological changes.

Allopolyploidy

Polyplody can also result from hybridizing two closely related species. If a haploid ovum from a species with chromosome sets AA is fertilized by sperm from a species with sets BB , the resulting hybrid is AB , where $A = a_1, a_2, a_3, \dots, a_n$ and $B = b_1, b_2, b_3, \dots, b_n$. The hybrid plant may be sterile because of its inability to produce

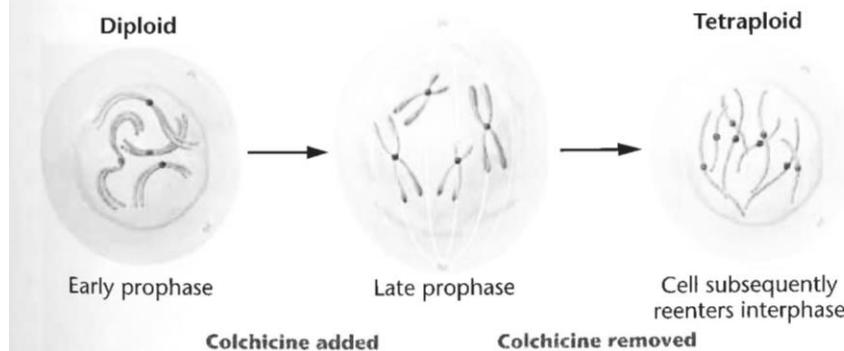


FIGURE 6.7 The potential involvement of colchicine in doubling the chromosome number. Two pairs of homologous chromosomes are shown. While each chromosome had replicated its DNA earlier during interphase, the chromosomes do not appear as double structures until late prophase. When anaphase fails to occur normally, the chromosome number doubles if the cell reenters interphase.

viable gametes. Most often, this occurs when some or all of the *a* and *b* chromosomes are not homologous and therefore cannot synapse in meiosis. As a result, unbalanced genetic conditions result. If, however, the new *AB* genetic combination undergoes a natural or an induced chromosomal doubling, two copies of all *a* chromosomes and two copies of all *b* chromosomes will be present, and they will pair during meiosis. As a result, a fertile *AABB* tetraploid is produced. These events are shown in Figure 6.8. Since this polyploid contains the equivalent of four haploid genomes derived from separate species, such an organism is called an **allotetraploid**. When both original species are known, an equivalent term, **amphidiploid**, is preferred in describing the allotetraploid.

Amphidiploid plants are often found in nature. Their reproductive success is based on their potential for forming balanced gametes. Since two homologs of each specific chromosome are present, meiosis occurs normally (Figure 6.8) and fertilization successfully propagates

the plant sexually. This discussion assumes the simplest situation, where none of the chromosomes in set *A* are homologous to those in set *B*. In amphidiploids, formed from closely related species, some homology between *a* and *b* chromosomes is likely. Allopolyploids are rare in most animals because mating behavior is most often species-specific, and thus the initial step in hybridization is unlikely to occur.

A classical example of amphidiploidy in plants is the cultivated species of American cotton, *Gossypium* (Figure 6.9). This species has 26 pairs of chromosomes: 13 are large and 13 are much smaller. When it was discovered that Old World cotton had only 13 pairs of large chromosomes, allopolyploidy was suspected. After an examination of wild American cotton revealed 13 pairs of small chromosomes, this speculation was strengthened. J. O. Beasley reconstructed the origin of cultivated cotton experimentally by crossing the Old World strain with the wild American strain and then treating the hybrid with colchicine to double the chromosome number. The result of these treatments was a fertile amphidiploid variety of cotton. It contained 26 pairs of chromosomes as well as characteristics similar to the cultivated variety.

Amphidiploids often exhibit traits of both parental species. An interesting example, but one with no practical economic importance, is that of the hybrid formed between the radish *Raphanus sativus* and the cabbage *Brassica oleracea*. Both species have a haploid number $n = 9$. The initial hybrid consists of nine *Raphanus* and nine *Brassica* chromosomes (9R + 9B). Although hybrids are almost always sterile, some fertile amphidiploids (18R + 18B) have been produced. Unfortunately, the root of this plant is more like the cabbage and its shoot is more like the radish; had the converse occurred, the hybrid might have been of economic importance.

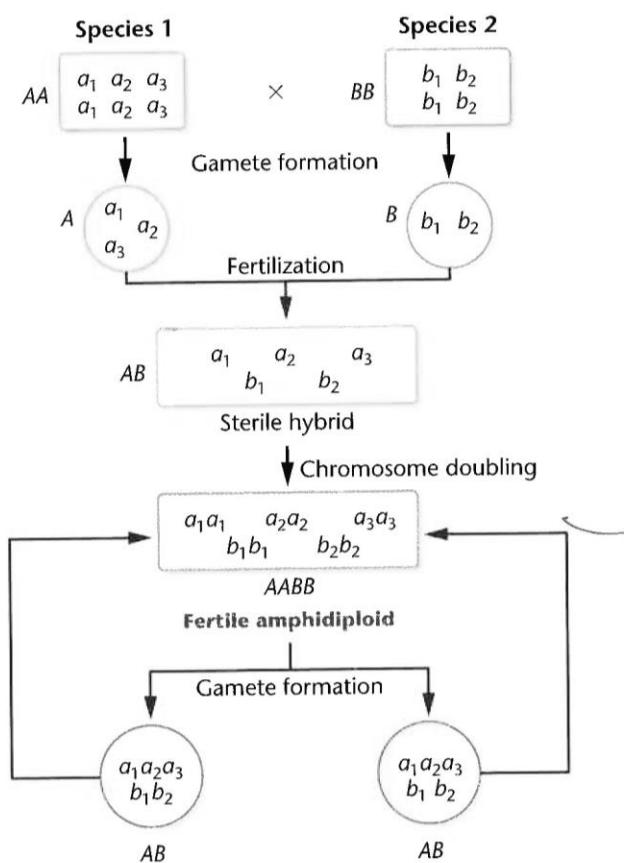


FIGURE 6.8 The origin and propagation of an amphidiploid. Species 1 contains genome *A* consisting of three distinct chromosomes, a_1, a_2 , and a_3 . Species 2 contains genome *B* consisting of two distinct chromosomes, b_1 and b_2 . Following fertilization between members of the two species and chromosome doubling, a fertile amphidiploid containing two complete diploid genomes (*AABB*) is formed.



FIGURE 6.9 The pods of the amphidiploid form of *Gossypium*, the cultivated American cotton plant.

A much more successful commercial hybridization uses the grasses wheat and rye. Wheat (genus *Triticum*) has a basic haploid genome of seven chromosomes. In addition to normal diploids ($2n = 14$), cultivated autopolyploids exist, including tetraploid ($4n = 28$) and hexaploid ($6n = 42$) species. Rye (genus *Secale*) also has a genome consisting of seven chromosomes. The only cultivated species is the diploid plant ($2n = 14$).

Using the technique outlined in Figure 6.8, geneticists have produced various hybrids. When tetraploid wheat is crossed with diploid rye and the F_1 plants are treated with colchicine, a hexaploid variety ($6n = 42$) is obtained; the hybrid, designated *Triticale*, represents a new genus. Fertile hybrid varieties derived from various wheat and rye species can be crossed or backcrossed. These crosses have created many variations of the genus *Triticale*. The hybrid plants demonstrate characteristics of both wheat and rye. For example, certain hybrids combine the high-protein content of wheat with rye's high content of the amino acid lysine. (The lysine content is low in wheat and thus is a limiting nutritional factor.) Wheat is considered to be a high-yielding grain, whereas rye is noted for its versatility of growth in unfavorable environments. *Triticale* species, which combine both traits, have the potential of significantly increasing grain production. Programs designed to improve crops through hybridization have long been under way in several developing countries.

Endopolyploidy

Endopolyploidy is the condition in which only certain cells in an otherwise diploid organism are polyploid. In such cells, the set of chromosomes replicates repeatedly without nuclear division. Numerous examples of naturally occurring endopolyploidy have been observed. For example, vertebrate liver cell nuclei, including those in humans, often contain $4n$, $8n$, or $16n$ chromosome sets. The stem and parenchymal tissue of apical regions of flowering plants are also often endopolyploid. Cells lining the gut of mosquito larvae attain a $16n$ ploidy, but during the pupal stages, such cells undergo very quick reduction divisions, giving rise to smaller diploid cells. In the water strider *Gerris*, wide variations in chromosome numbers are found in different tissues, with as many as 1024 to 2048 copies of each chromosome in the salivary gland cells. Since the diploid number in this organism is 22, the nuclei of these cells may contain more than 40,000 chromosomes.

Although the role of endopolyploidy is not clear, the proliferation of chromosome copies often occurs in cells where high levels of certain gene products are required.

1.2.7

Klug et al. (2019)

Page 378-380

15



Gene Mutation, DNA Repair, and Transposition

Pigment mutations within an ear of corn, caused by transposition of the *Ds* element.

CHAPTER CONCEPTS

- Mutations comprise any change in the nucleotide sequence of an organism's genome.
- Mutations are a source of genetic variation and provide the raw material for natural selection. They are also the source of genetic damage that contributes to cell death, genetic diseases, and cancer.
- Mutations have a wide range of effects on organisms depending on the type of base-pair alteration, the location of the mutation within the chromosome, and the function of the affected gene product.
- Mutations can occur spontaneously as a result of natural biological and chemical processes, or they can be induced by external factors, such as chemicals or radiation.
- Single-gene mutations cause a wide variety of human diseases.
- Organisms rely on a number of DNA repair mechanisms to counteract mutations. These mechanisms range from proofreading and correction of replication errors to base excision and homologous recombination repair.
- Mutations in genes whose products control DNA repair lead to genome hypermutability, human DNA repair diseases, and cancers.
- Transposable elements may move into and out of chromosomes, causing chromosome breaks and inducing mutations both within coding regions and in gene-regulatory regions.

The ability of DNA molecules to store, replicate, transmit, and decode information is the basis of genetic function. But equally important are the changes that occur to DNA sequences. Without the variation that arises from changes in DNA sequences, there would be no phenotypic variability, no adaptation to environmental changes, and no evolution. Gene mutations are the source of new alleles and are the origin of genetic variation within populations. On the downside, they are also the source of genetic changes that can lead to cell death, genetic diseases, and cancer.

Mutations also provide the basis for genetic analysis. The phenotypic variations resulting from mutations allow geneticists to identify and study the genes responsible for the modified trait. In genetic investigations, mutations act as identifying “markers” for genes so that they can be followed during their transmission from parents to offspring. Without phenotypic variability, classical genetic analysis would be impossible. For example, if all pea plants displayed a uniform phenotype, Mendel would have had no foundation for his research.

We have examined mutations in large regions of chromosomes—chromosomal mutations (see Chapter 6). In contrast, the mutations we will now explore are those occurring primarily in the base-pair sequence of

DNA within and surrounding individual genes—**gene mutations**. We will also describe how the cell defends itself from mutations using various mechanisms of DNA repair.

15.1 Gene Mutations Are Classified in Various Ways

A mutation can be defined as an alteration in the nucleotide sequence of an organism's genome. Any base-pair change in any part of a DNA molecule can be considered a mutation. A mutation may comprise a single base-pair substitution, a deletion or insertion of one or more base pairs, or a major alteration in the structure of a chromosome. The genomes of RNA viruses are made up of single-stranded or double-stranded RNA molecules. These RNA-based genomes are also subject to changes in ribonucleotide sequence that result in mutations. In this chapter, we will restrict our discussion to mutations that occur in DNA genomes.

Mutations may occur within regions of a gene that code for protein or within noncoding regions of a gene such as introns and regulatory sequences, including promoters, enhancers, and splicing signals. Mutations may or may not bring about a detectable change in phenotype. The extent to which a mutation changes the characteristics of an organism depends on which type of cell suffers the mutation and the degree to which the mutation alters the function of a gene product or a gene-regulatory region.

Mutations can occur in somatic cells or within germ cells. Those that occur in germ cells are heritable and are the basis for the transmission of genetic diversity and evolution, as well as genetic diseases. Those that occur in somatic cells are not transmitted to the next generation but may lead to altered cellular function or tumors.

Because of the wide range of types and effects of mutations, geneticists classify mutations according to several

different schemes. These organizational schemes are not mutually exclusive. In this section, we outline some of the ways in which gene mutations are classified.

Classification Based on Type of Molecular Change

Geneticists often classify gene mutations in terms of the nucleotide changes that constitute the mutation. A change of one base pair to another in a DNA molecule is known as a **point mutation**, or **base substitution** (see Figure 15.1). A change of one nucleotide of a triplet within a protein-coding portion of a gene may result in the creation of a new triplet that codes for a different amino acid in the protein product. If this occurs, the mutation is known as a **missense mutation**. A second possible outcome is that the triplet will be changed into a stop codon, resulting in the termination of translation of the protein. This is known as a **nonsense mutation**. If the point mutation alters a codon but does not result in a change in the amino acid at that position in the protein (due to degeneracy of the genetic code), it can be considered a **silent mutation**.

Because eukaryotic genomes consist of so much more noncoding DNA than coding DNA (see Chapter 12), the vast majority of mutations are likely to occur in noncoding regions. These mutations may be considered **neutral mutations** if they do not affect gene products or gene expression. Most silent mutations, which do not change the amino acid sequence of the encoded protein, can also be considered neutral mutations. However, some silent mutations may alter a DNA sequence that codes for regulatory function, such as an RNA splicing signal, resulting in an altered protein and a discernible phenotype.

You will often see two other terms used to describe base substitutions. If a pyrimidine replaces a pyrimidine or a purine replaces a purine, a **transition** has occurred. If a purine replaces a pyrimidine, or vice versa, a **transversion** has occurred.

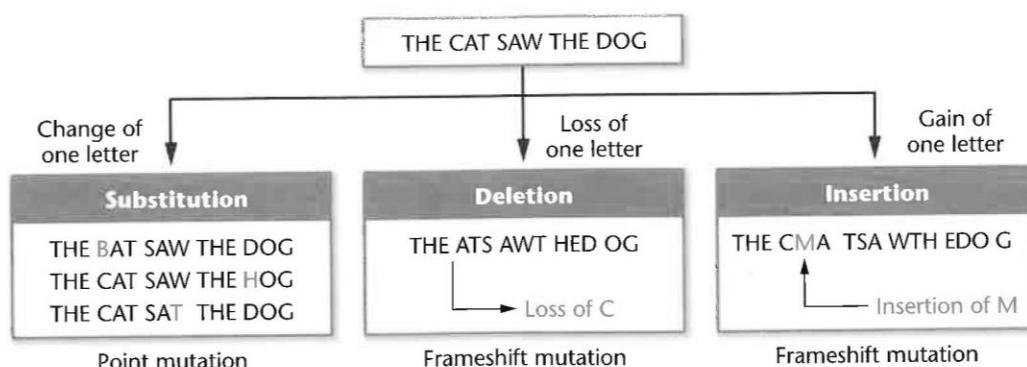


FIGURE 15.1 Analogy showing the effects of substitution, deletion, and insertion of one letter in a sentence composed of three-letter words to demonstrate point and frameshift mutations.

Another type of change is the insertion or deletion of one or more nucleotides at any point within the gene. As illustrated in Figure 15.1, the loss or addition of a single nucleotide causes all of the subsequent three-letter codons to be changed. These are called **frameshift mutations** because the frame of triplet reading during translation is altered. A frameshift mutation will occur when any number of bases are added or deleted, except multiples of three, which would reestablish the initial frame of reading (see Figure 13.2). It is possible that one of the many altered triplets will be UAA, UAG, or UGA, the translation termination codons. When one of these triplets is encountered during translation, polypeptide synthesis is terminated at that point. Obviously, the results of frameshift mutations can be very severe, such as producing a truncated protein or defective enzymes, especially if they occur early in the coding sequence.

Classification Based on Effect on Function

As discussed earlier (see Chapter 4), a **loss-of-function mutation** is one that reduces or eliminates the function of the gene product. Mutations that result in complete loss of function are known as **null mutations**. Any type of mutation, from a point mutation to deletion of the entire gene, may lead to a loss of function.

Most loss-of-function mutations are recessive. A **recessive mutation** results in a wild-type phenotype when present in a diploid organism and the other allele is wild type. In this case, the presence of less than 100 percent of the gene product is sufficient to bring about the wild-type phenotype.

Some loss-of-function mutations can be dominant. A **dominant mutation** results in a mutant phenotype in a diploid organism, even when the wild-type allele is also present. Dominant mutations in diploid organisms can have several different types of effects. A **dominant negative mutation** in one allele may encode a gene product that is inactive and directly interferes with the function of the product of the wild-type allele. For example, this can occur when the nonfunctional gene product binds to the wild-type gene product in a homodimer, inactivating or reducing the activity of the homodimer.

A dominant negative mutation can also result from **haploinsufficiency**, which occurs when one allele is inactivated by mutation, leaving the individual with only one functional copy of a gene. The active allele may be a wild-type copy of the gene but does not produce enough wild-type gene product to bring about a wild-type phenotype. In humans, Marfan syndrome is an example of a disorder caused by haploinsufficiency—in this case as a result of a loss-of-function mutation in one copy of the *fibrillin-1* (*FBN1*) gene.

In contrast, a **gain-of-function mutation** codes for a gene product with enhanced, negative, or new functions. This may be due to a change in the amino acid sequence of the protein that confers a new activity, or it may result from a mutation in a regulatory region of the gene, leading to expression of the gene at higher levels or at abnormal times or places. Typically, gain-of-function mutations are dominant.

A **suppressor mutation** is a second mutation that either reverts or relieves the effects of a previous mutation. A suppressor mutation can occur within the same gene that suffered the first mutation (**intragenic mutation**) or elsewhere in the genome (**intergenic mutation**).

An example of an intragenic suppressor mutation is one that reverts a frameshift mutation. For instance, if the original frameshift mutation was caused by a *deletion* of one or two base pairs within a gene's reading frame, the reading frame would be altered, as illustrated in Figure 15.1. However, if a second mutation occurs near the first frameshift mutation, and it involves the *insertion* of one or two base pairs, the original reading frame of the gene may be restored. An incorrect reading frame would still exist for a short distance near the mutations, but the resulting protein would be primarily wild type and could be functional. (The ways in which intragenic suppressor mutations were used experimentally to determine the triplet reading frame of DNA were described in Chapter 13.)

Another example of an intragenic suppressor mutation is one that creates a codon specifying a correct (or similar) amino acid, so as to restore function to a mutated gene product. For instance, if the first mutation changed the sequence 5'-TTA-3' (which codes for leucine) to 5'-GTA-3' (which codes for valine), then a second mutation occurring in the valine codon, changing it to 5'-CTA-3', would restore the codon to one that codes for leucine.

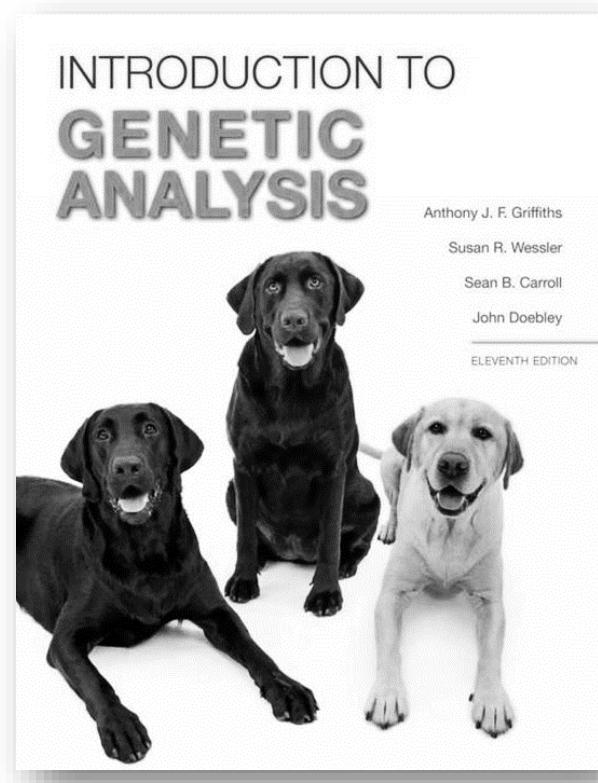
An example of a mutation that would act as an intergenic suppressor mutation would be as follows. A mutated gene may encode a protein whose structure has been altered so that it will not interact with another protein with which it would normally interact. If the gene encoding the second protein acquires a mutation that alters the structure of its gene product in such a way that it can now interact with the first mutant protein, the second mutation would be considered an intergenic suppressor mutation.

Depending on their type and location, mutations can have a wide range of phenotypic effects, from none to severe. Some examples of mutation types based on their phenotypic outcomes are listed in **Table 15.1**.

Classification Based on Location of Mutation

Mutations may be classified according to the cell type or chromosomal locations in which they occur. **Somatic**

1.3 Study book



Title:	<i>Introduction to Genetic Analysis</i>
Edition:	11
Authors:	<i>Anthony J.F. Griffiths, Susan R. Wessler, Sean B. Carroll, and Richard C. Lewontin</i>
Year:	2015
Publisher:	<i>Macmillan Learning</i>

1.3.1

Griffiths et al. (2015)

Page 58-60

The F₂ consists of one-half red-eyed and one-half white-eyed flies of both sexes. Hence, in sex linkage, we see examples not only of different ratios in different sexes, but also of differences between reciprocal crosses.

Note that *Drosophila* eye color has nothing to do with sex determination, and so we have an illustration of the principle that genes on the sex chromosomes are not necessarily related to sexual function. The same is true in humans: in the discussion of pedigree analysis later in this chapter, we shall see many X-linked genes, yet few could be construed as being connected to sexual function.

The abnormal allele associated with white eye color in *Drosophila* is recessive, but abnormal alleles of genes on the X chromosome that are dominant also arise, such as the *Drosophila* mutant hairy wing (*Hw*). In such cases, the wild-type allele (*Hw*⁺) is recessive. The dominant abnormal alleles show the inheritance pattern corresponding to that of the wild-type allele for red eyes in the preceding example. The ratios obtained are the same.

KEY CONCEPT Sex-linked inheritance regularly shows different phenotypic ratios in the two sexes of progeny, as well as different ratios in reciprocal crosses.

Historically, in the early decades of the twentieth century, the demonstration by Morgan of X-linked inheritance of *white eyes* in *Drosophila* was a key piece of evidence that suggested that genes are indeed located on chromosomes, because an inheritance pattern was correlated with one specific chromosome pair. The idea became known as “the chromosome theory of inheritance.” At that period in history, it had recently been shown that, in many organisms, sex is determined by an X and a Y chromosome and that, in males, these chromosomes segregate equally at meiosis to regenerate equal numbers of males and females in the next generation. Morgan recognized that the inheritance of alleles of the eye-color gene is exactly parallel to the inheritance of X chromosomes at meiosis; hence, the gene was likely to be on the X chromosome. The inheritance of *white eyes* was extended to *Drosophila* lines that had abnormal numbers of sex chromosomes. With the use of this novel situation, it was still possible to predict gene-inheritance patterns from the segregation of the abnormal chromosomes. That these predictions proved correct was a convincing test of the chromosome theory.

Other genetic analyses revealed that, in chickens and moths, sex-linked inheritance could be explained only if the female was the heterogametic sex. In these organisms, the female sex chromosomes were designated ZW and males were designated ZZ.

2.6 Human Pedigree Analysis

Human matings, like those of experimental organisms, provide many examples of single-gene inheritance. However, controlled experimental crosses cannot be made with humans, and so geneticists must resort to scrutinizing medical records in the hope that informative matings have been made (such as monohybrid crosses) that could be used to infer single-gene inheritance. Such a scrutiny of records of matings is called **pedigree analysis**. A member of a family who first comes to the attention of a geneticist is called the **propositus**. Usually, the phenotype of the propositus is exceptional in some way; for example, the propositus might have some type of medical disorder. The investigator then traces the history of the phenotype through the history of the family and draws a family tree, or pedigree, by using the standard symbols given in Figure 2-18.

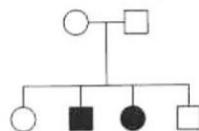
To see single-gene inheritance, the patterns in the pedigree have to be interpreted according to Mendel’s law of equal segregation, but humans usually have few children and so, because of this small progeny sample size, the expected

3:1 and 1:1 ratios are usually not seen unless many similar pedigrees are combined. The approach to pedigree analysis also depends on whether one of the contrasting phenotypes is a rare disorder or both phenotypes of a pair are common (in which case they are said to be “morphs” of a polymorphism). Most pedigrees are drawn for medical reasons and therefore concern medical disorders that are almost by definition rare. In this case, we have two phenotypes: the presence and the absence of the disorder. Four patterns of single-gene inheritance are revealed in pedigrees. Let’s look, first, at recessive disorders caused by recessive alleles of single autosomal genes.

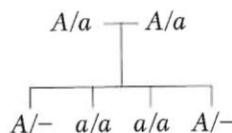
Autosomal recessive disorders

The affected phenotype of an autosomal recessive disorder is inherited as a recessive allele; hence, the corresponding unaffected phenotype must be inherited as the corresponding dominant allele. For example, the human disease phenylketonuria (PKU), discussed earlier, is inherited in a simple Mendelian manner as a recessive phenotype, with PKU determined by the allele *p* and the normal condition determined by *P*. Therefore, people with this disease are of genotype *p/p*, and people who do not have the disease are either *P/P* or *P/p*. Recall that the term *wild type* and its allele symbols are not used in human genetics because wild type is impossible to define.

What patterns in a pedigree would reveal autosomal recessive inheritance? The two key points are that (1) generally the disorder appears in the progeny of unaffected parents and (2) the affected progeny include both males and females. When we know that both male and female progeny are affected, we can infer that we are most likely dealing with simple Mendelian inheritance of a gene on an autosome, rather than a gene on a sex chromosome. The following typical pedigree illustrates the key point that affected children are born to unaffected parents:



From this pattern, we can deduce a simple monohybrid cross, with the recessive allele responsible for the exceptional phenotype (indicated in black). Both parents must be heterozygotes—say, *A/a*; both must have an *a* allele because each contributed an *a* allele to each affected child, and both must have an *A* allele because they are phenotypically normal. We can identify the genotypes of the children (shown left to right) as *A/-*, *a/a*, *a/a*, and *A/-*. Hence, the pedigree can be rewritten as follows:



This pedigree does not support the hypothesis of X-linked recessive inheritance, because, under that hypothesis, an affected daughter must have a heterozygous mother (possible) and a hemizygous father, which is clearly

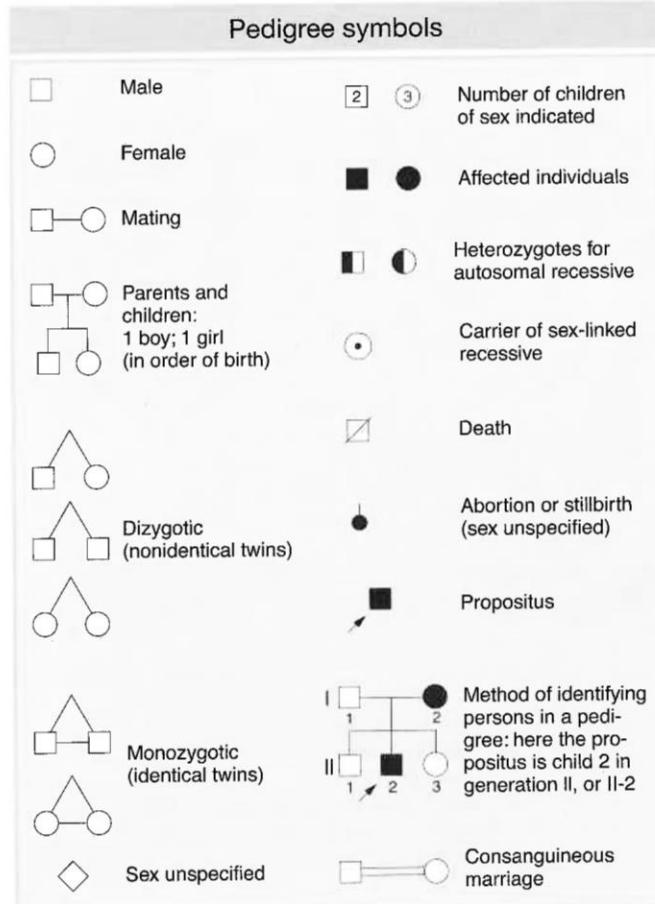


FIGURE 2-18 A variety of symbols are used in human pedigree analysis.

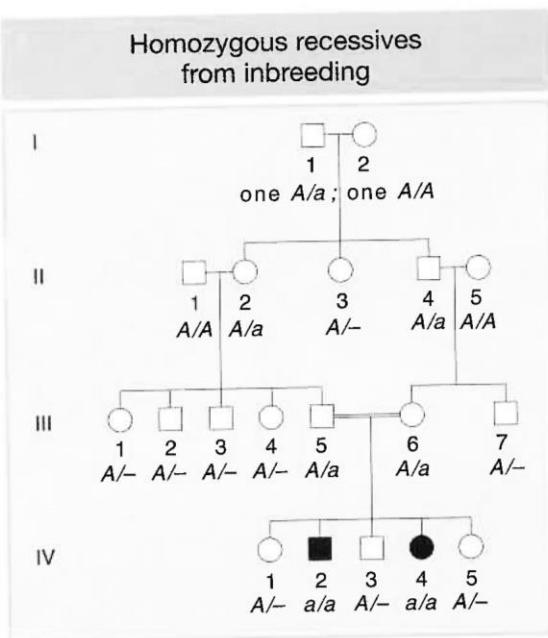


FIGURE 2-19 Pedigree of a rare recessive phenotype determined by a recessive allele *a*. Gene symbols are normally not included in pedigree charts, but genotypes are inserted here for reference. Persons II-1 and II-5 marry into the family; they are assumed to be normal because the heritable condition under scrutiny is rare. Note also that it is not possible to be certain of the genotype in some persons with normal phenotype; such persons are indicated by *A/-*. Persons III-5 and III-6, who generate the recessives in generation IV, are first cousins. They both obtain their recessive allele from a grandparent, either I-1 or I-2.

impossible because the father would have expressed the phenotype of the disorder.

Notice that, even though Mendelian rules are at work, Mendelian ratios are not necessarily observed in single families because of small sample size, as predicted earlier. In the preceding example, we observe a 1:1 phenotypic ratio in the progeny of a monohybrid cross. If the couple were to have, say, 20 children, the ratio would be something like 15 unaffected children and 5 with PKU (a 3:1 ratio), but, in a small sample of 4 children, any ratio is possible, and all ratios are commonly found.

The family pedigrees of autosomal recessive disorders tend to look rather bare, with few black symbols. A recessive condition shows up in groups of affected siblings, and the people in earlier and later generations tend not to be affected. To understand why this is so, it is important to have some understanding of the genetic structure of populations underlying such rare conditions. By definition, if the condition is rare, most people do not carry the abnormal allele. Furthermore, most of those people who do carry the abnormal allele are heterozygous for it rather than homozygous. The basic reason that heterozygotes are much more common than recessive homozygotes is that, to be a recessive homozygote, both parents must have the *a* allele, but, to be a heterozygote, only one parent must have it.

The birth of an affected person usually depends on the rare chance union of unrelated heterozygous parents. However, inbreeding (mating between relatives, sometimes referred to as *consanguinity* in humans) increases the chance that two heterozygotes will mate. An example of a marriage between cousins is shown in Figure 2-19. Individuals III-5 and III-6 are first cousins and produce two homozygotes for the rare allele. You can see from Figure 2-19 that an ancestor who is a heterozygote may produce many descendants who also are heterozygotes. Hence, two cousins can carry the same rare recessive allele inherited from a common ancestor. For two *unrelated* persons to be heterozygous, they would have to inherit the rare allele from *both* their families. Thus, matings between relatives generally run a higher risk of producing recessive disorders than do matings between non-relatives. For this reason, first-cousin marriages contribute a large proportion of people with recessive diseases in the population.

Some other examples of human recessive disorders are shown in Figure 2-20. Cystic fibrosis is a disease inherited on chromosome 7 according to Mendelian rules as an autosomal recessive phenotype. Its most important symptom is the secretion of large amounts of mucus into the lungs, resulting in death from a combination of effects but usually precipitated by infection of the respiratory tract. The mucus can be dislodged by mechanical chest thumpers, and pulmonary infection can be prevented by antibiotics; thus, with treatment, cystic fibrosis patients can live to adulthood. The cystic fibrosis gene (and its mutant allele) was one of the first human disease genes to be isolated at the DNA level, in 1989. This line of research eventually revealed that the disorder is caused by a defective protein that normally transports chloride ions across the cell membrane. The resultant alteration of the salt balance changes the constitution of the lung mucus. This new understanding of gene function in affected and unaffected persons has given hope for more effective treatment.

Human albinism also is inherited in the standard autosomal recessive manner. The mutant allele is of a gene that normally synthesizes the brown or black pigment melanin, normally found in skin, hair, and the retina of the eye (Figure 2-21).

KEY CONCEPT In human pedigrees, an autosomal recessive disorder is generally revealed by the appearance of the disorder in the male and female progeny of unaffected parents.

1.3.2

Griffiths et al. (2015)

Page 128-142

Some of the questions that geneticists want to answer about the genome are, *What genes are present in the genome? What functions do they have? What positions do they occupy on the chromosomes?* Their pursuit of the third question is broadly called mapping. Mapping is the main focus of this chapter, but all three questions are interrelated, as we will see.

We all have an everyday feeling for the importance of maps in general, and, indeed, we have all used them at some time in our lives to find our way around. Relevant to the focus of this chapter is that, in some situations, several maps need to be used simultaneously. A good example in everyday life is in navigating the dense array of streets and buildings of a city such as London, England. A street map that shows the general layout is one necessity. However, the street map is used by tourists and Londoners alike in conjunction with another map, that of the underground railway system. The underground system is so complex and spaghetti-like that, in 1933, an electrical circuit engineer named Harry Beck drew up the streamlined (although distorted) map that has remained to this day an icon of London. The street and underground maps of London are compared in Figure 4-1. Note that the positions of the underground stations and the exact distances between them are of no interest in themselves, except as a way of getting to a destination of interest such as Westminster Abbey. We will see three parallels with the London maps when chromosome maps are used to zero in on individual “destinations,” or specific genes. First, several different types of chromosome maps are often necessary and must be used in conjunction; second, maps that contain distortions are still useful; and third, many sites on a chromosome map are

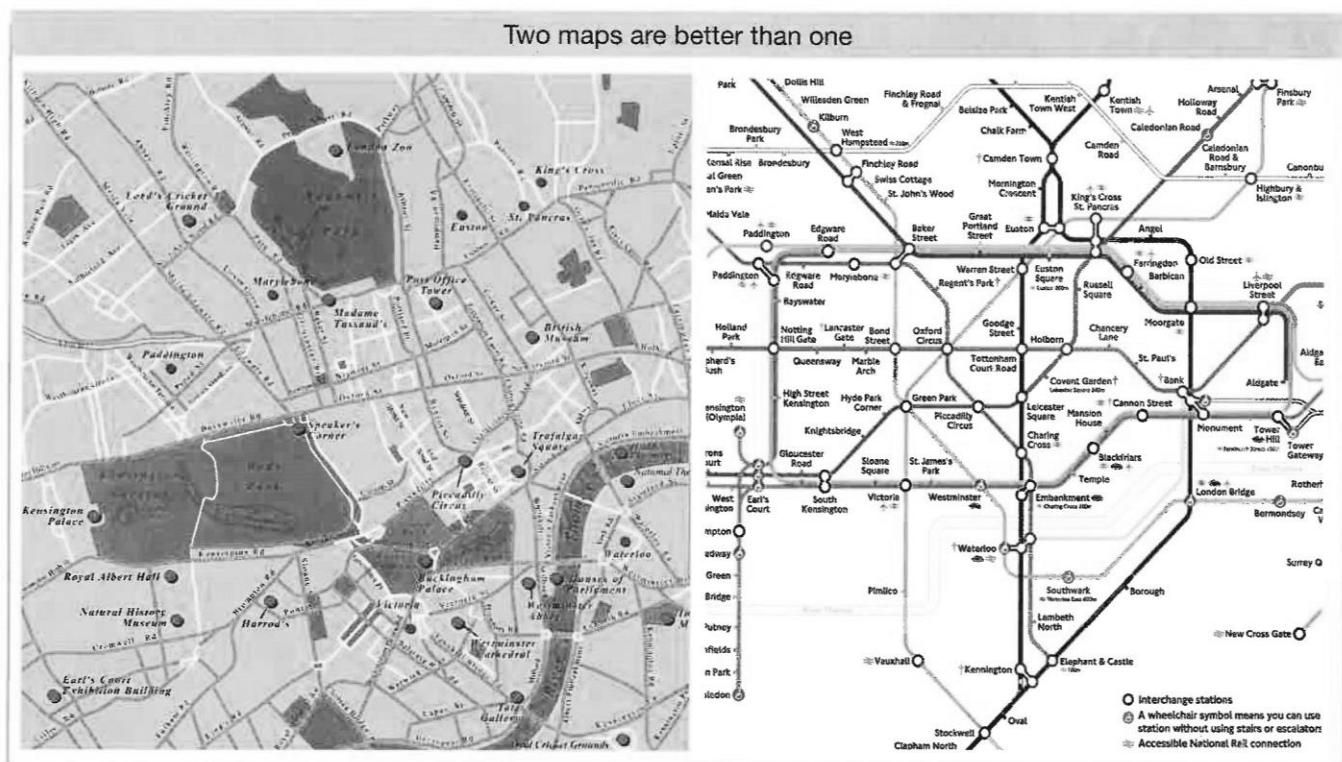


FIGURE 4-1 These London maps illustrate the principle that, often, several maps are needed to get to a destination of interest. The map of the underground railway (“the Tube”) is used to get to a destination of interest such as a street address, shown on the street map. In genetics, two different kinds of genome maps are often useful in locating a gene, leading to an understanding its structure and function. [(left) © MAPS.com/Corbis; (right) Transport for London.]

charted only because they are useful in trying to zero in on other sites that are the ones of real interest.

Obtaining a map of gene positions on the chromosomes is an endeavor that has occupied thousands of geneticists for the past 80 years or so. Why is it so important? There are several reasons:

1. Gene position is crucial information needed to build *complex genotypes* required for experimental purposes or for commercial applications. For example, in Chapter 6, we will see cases in which special allelic combinations must be put together to explore gene interaction.
2. Knowing the position occupied by a gene provides a way of discovering its *structure and function*. A gene's position can be used to define it at the DNA level. In turn, the DNA sequence of a wild-type gene or its mutant allele is a necessary part of deducing its underlying function.
3. The genes present and their arrangement on chromosomes are often slightly different in related species. For example, the rather long human chromosome number 2 is split into two shorter chromosomes in the great apes. By comparing such differences, geneticists can deduce the *evolutionary genetic mechanisms* through which these genomes diverged. Hence, chromosome maps are useful in interpreting mechanisms of evolution.

The arrangement of genes on chromosomes is represented diagrammatically as a unidimensional **chromosome map**, showing gene positions, known as **loci** (sing., locus), and the distances between the loci based on some kind of scale. Two basic types of chromosome maps are currently used in genetics; they are assembled by quite different procedures yet are used in a complementary way. *Recombination-based maps*, which are the topic of this chapter, map the loci of genes that have been identified by mutant phenotypes showing single-gene inheritance. *Physical maps* (see Chapter 14) show the genes as segments arranged along the long DNA molecule that constitutes a chromosome. These maps show different views of the genome, but, just like the maps of London, they can be used together to arrive at an understanding of what a gene's function is at the molecular level and how that function influences phenotype.

KEY CONCEPT Genetic maps are useful for strain building, for interpreting evolutionary mechanisms, and for discovering a gene's unknown function. Discovering a gene's function is facilitated by integrating information on recombination-based and physical maps.

4.1 Diagnostics of Linkage

Recombination maps of chromosomes are usually assembled two or three genes at a time, with the use of a method called linkage analysis. When geneticists say that two genes are **linked**, they mean that the loci of those genes are on the same chromosome, and, hence, the alleles on any one homolog are physically joined (linked) by the DNA between them. The way in which early geneticists deduced linkage is a useful means of introducing most of the key ideas and procedures in the analysis.

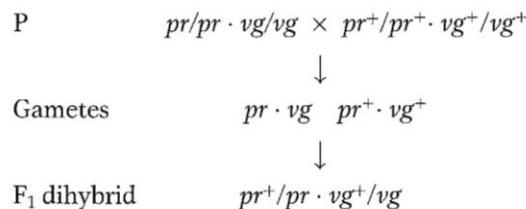
Using recombinant frequency to recognize linkage

In the early 1900s, William Bateson and R. C. Punnett (for whom the Punnett square was named) were studying the inheritance of two genes in sweet peas. In

a standard self of a dihybrid F_1 , the F_2 did not show the 9:3:3:1 ratio predicted by the principle of independent assortment. In fact, Bateson and Punnett noted that certain combinations of alleles showed up more often than expected, almost as though they were physically attached in some way. However, they had no explanation for this discovery.

Later, Thomas Hunt Morgan found a similar deviation from Mendel's second law while studying two autosomal genes in *Drosophila*. Morgan proposed linkage as a hypothesis to explain the phenomenon of apparent allele association.

Let's look at some of Morgan's data. One of the genes affected eye color (pr , purple, and pr^+ , red), and the other gene affected wing length (vg , vestigial, and vg^+ , normal). (Vestigial wings are very small compared to wild type.) The wild-type alleles of both genes are dominant. Morgan performed a cross to obtain dihybrids and then followed with a testcross:



Testcross:

$$\begin{array}{ccc}
 pr^+/pr \cdot vg^+/vg \text{ ♀} & \times & pr/pr \cdot vg/vg \text{ ♂} \\
 \text{F}_1 \text{ dihybrid female} & & \text{Tester male}
 \end{array}$$

Morgan's use of the testcross is important. Because the tester parent contributes gametes carrying only recessive alleles, the phenotypes of the offspring directly reveal the alleles contributed by the gametes of the dihybrid parent, as described in Chapters 2 and 3. Hence, the analyst can concentrate on meiosis in one parent (the dihybrid) and essentially forget about meiosis in the other (the tester). In contrast, from an F_1 *self*, there are two sets of meioses to consider in the analysis of progeny: one in the male parent and the other in the female.

Morgan's testcross results were as follows (listed as the gametic classes from the dihybrid):

$pr^+ \cdot vg^+$	1339
$pr \cdot vg$	1195
$pr^+ \cdot vg$	151
$pr \cdot vg^+$	154
	2839

Obviously, these numbers deviate drastically from the Mendelian prediction of a 1:1:1:1 ratio expected from independent assortment (approximately 710 in each of the four classes). In Morgan's results, we see that the first two allele combinations are in the great majority, clearly indicating that they are associated, or "linked."

Another useful way of assessing the testcross results is by considering the percentage of recombinants in the progeny. By definition, the recombinants in the present cross are the two types $pr^+ \cdot vg$ and $pr \cdot vg^+$ because they are clearly not the two input genotypes contributed to the F_1 dihybrid by the original homozygous parental flies (more precisely, by their gametes). We see that the two recombinant types are approximately equal in frequency (151 ~ 154). Their total is 305, which is a frequency of $(305/2839) \times 100$, or 10.7 percent. We can make sense of

these data, as Morgan did, by postulating that the genes were linked on the same chromosome, and so the parental allelic combinations are held together in the majority of progeny. In the dihybrid, the allelic conformation must have been as follows:

$$\begin{array}{c} pr^+ \quad vg^+ \\ \hline pr \quad vg \end{array}$$

The tendency of linked alleles to be inherited as a package is illustrated in Figure 4-2.

Now let's look at another cross that Morgan made with the use of the same alleles but in a different combination. In this cross, each parent is homozygous for the wild-type allele of one gene and the mutant allele of the other. Again, F₁ females were testcrossed:

P	$pr^+/pr^+ \cdot vg/vg \times pr/pr \cdot vg^+/vg^+$
	↓
Gametes	$pr^+ \cdot vg \quad pr \cdot vg^+$
	↓
F ₁ dihybrid	$pr^+/pr \cdot vg^+/vg$

Testcross:

$$\begin{array}{ccc} pr^+/pr \cdot vg^+/vg & \times & pr/pr \cdot vg/vg \\ \text{♀} & & \delta \\ \text{F}_1 \text{ dihybrid female} & & \text{Tester male} \end{array}$$

The following progeny were obtained from the testcross:

$pr^+ \cdot vg^+$	157
$pr \cdot vg$	146
$pr^+ \cdot vg$	965
$pr \cdot vg^+$	1067
	2335

Again, these results are not even close to a 1:1:1:1 Mendelian ratio. Now, however, the recombinant classes are the converse of those in the first analysis, $pr^+ \cdot vg^+$ and $pr \cdot vg$. But notice that their frequency is approximately the same: $(157 + 146)/2335 \times 100 = 12.9$ percent. Again, linkage is suggested, but, in this case, the F₁ dihybrid must have been as follows:

$$\begin{array}{c} pr^+ \quad vg \\ \hline pr \quad vg^+ \end{array}$$

Dihybrid testcross results like those just presented are commonly encountered in genetics. They follow the general pattern:

**Two equally frequent nonrecombinant classes totaling
in excess of 50 percent**

**Two equally frequent recombinant classes totaling
less than 50 percent**

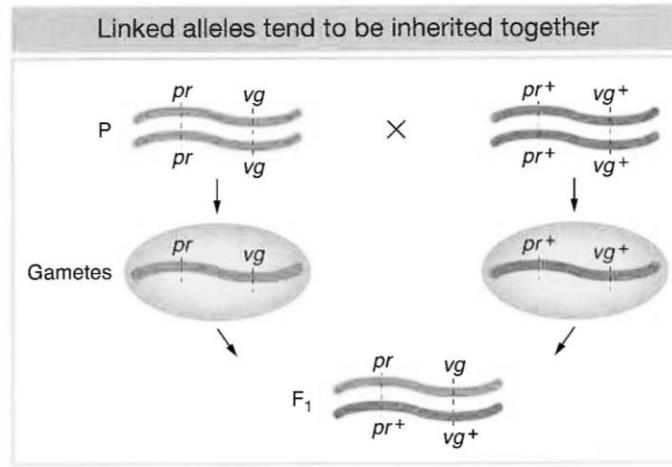


FIGURE 4-2 Simple inheritance of two genes located on the same chromosome pair. The same genes are present together on a chromosome in both parents and progeny.

KEY CONCEPT When two genes are close together on the same chromosome pair (that is, they are linked), they do not assort independently but produce a recombinant frequency of less than 50 percent. Hence, conversely, a recombinant frequency of less than 50 percent is a diagnostic for linkage.

Crossing over produces new allelic combinations

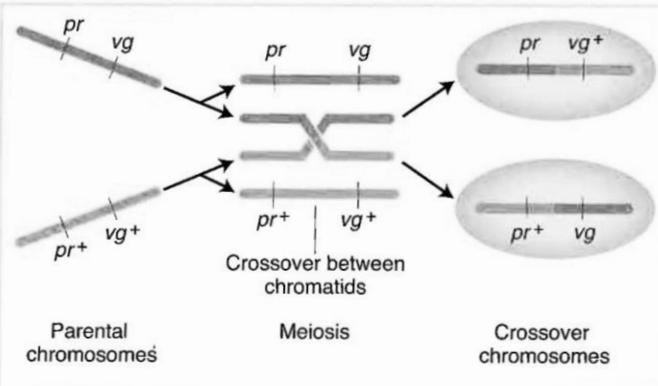


FIGURE 4-3 The exchange of parts by crossing over may produce gametic chromosomes whose allelic combinations differ from the parental combinations.

How crossovers produce recombinants for linked genes

The linkage hypothesis explains why allele combinations from the parental generations remain together: the genes are physically attached by the segment of chromosome between them. But exactly how are *any* recombinants produced when genes are linked? Morgan suggested that, when homologous chromosomes pair at meiosis, the chromosomes occasionally break and exchange parts in a process called **crossing over**. Figure 4-3 illustrates this physical exchange of chromosome segments. The two new combinations are called **crossover products**.

Is there any microscopically observable process that could account for crossing over? At meiosis, when duplicated homologous chromosomes pair with each other—in genetic terms, when the two dyads unite as a bivalent—a cross-shaped structure called a *chiasma* (pl., *chiasmata*) often forms between two nonsister chromatids. Chiasmata are shown in Figure 4-4. To Morgan, the appearance of the chiasmata visually corroborated the concept of crossing over. (Note that the chiasmata seem to indicate that *chromatids*, not unduplicated chromosomes, participate in a crossover. We will return to this point later.)

KEY CONCEPT For linked genes, recombinants are produced by crossovers. Chiasmata are the visible manifestations of crossovers.

Chiasmata are the sites of crossing over

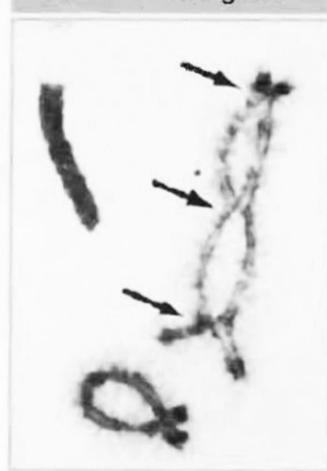


FIGURE 4-4 Several chiasmata appear in this photograph taken in the course of meiosis in a grasshopper testis. [G. H. Jones and F. C. H. Franklin, "Meiotic Crossing-over: Obligation and Interference," *Cell* 126:2 (28 July 2006), 246–248. © Elsevier.]

Linkage symbolism and terminology

The work of Morgan showed that linked genes in a dihybrid may be present in one of two basic conformations. In one, the two dominant, or wild-type, alleles are present on the same homolog (as in Figure 4-3); this arrangement is called a **cis conformation** (*cis* means "adjacent"). In the other, they are on different homologs, in what is called a **trans conformation** (*trans* means "opposite"). The two conformations are written as follows:

Cis	AB/ab or $++/ab$
Trans	Ab/aB or $+b/a+$

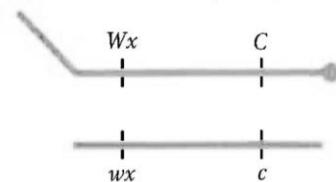
Note the following conventions that pertain to linkage symbolism:

1. Alleles on the same homolog have no punctuation between them.
2. A slash symbolically separates the two homologs.
3. Alleles are always written in the same order on each homolog.
4. As in earlier chapters, genes known to be on different chromosomes (unlinked genes) are shown separated by a semicolon—for example, $A/a; C/c$.
5. In this book, genes of *unknown* linkage are shown separated by a dot, $A/a \cdot D/d$.

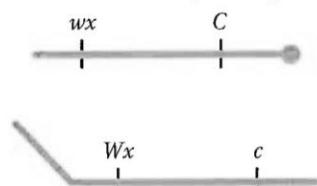
Evidence that crossing over is a breakage-and-rejoining process

The idea that recombinants are produced by some kind of exchange of material between homologous chromosomes was a compelling one. But experimentation was necessary to test this hypothesis. A first step was to find a case in which the exchange of parts between chromosomes would be visible under the microscope. Several investigators approached this problem in the same way, and one of their analyses follows.

In 1931, Harriet Creighton and Barbara McClintock were studying two genes of corn that they knew were both located on chromosome 9. One affected seed color (C , colored; c , colorless), and the other affected endosperm composition (Wx , waxy; wx , starchy). The plant was a dihybrid in *cis* conformation. However, in one plant, the chromosome 9 carrying the alleles C and Wx was unusual in that it also carried a large, densely staining element (called a *knob*) on the C end and a longer piece of chromosome on the Wx end; thus, the heterozygote was



In the progeny of a testcross of this plant, they compared recombinants and parental genotypes. They found that all the recombinants inherited one or the other of the two following chromosomes, depending on their recombinant makeup:



Thus, there was a precise correlation between the *genetic* event of the appearance of recombinants and the *chromosomal* event of crossing over. Consequently, the chiasmata appeared to be the sites of exchange, although what was considered to be the definitive test was not undertaken until 1978.

What can we say about the molecular mechanism of chromosome exchange in a crossover event? The short answer is that a crossover results from the breakage and reunion of DNA. Two parental chromosomes break at the same position, and then each piece joins up with the neighboring piece from the *other* chromosome. In Section 4.8, we will see a model of the molecular processes that allow DNA to break and rejoin in a precise manner such that no genetic material is lost or gained.

KEY CONCEPT A crossover is the breakage of two DNA molecules at the same position and their rejoining in two reciprocal recombinant combinations.

Evidence that crossing over takes place at the four-chromatid stage

As already noted, the diagrammatic representation of crossing over in Figure 4-3 shows a crossover taking place at the four-chromatid stage of meiosis; in other words, crossovers are between nonsister chromatids. However, it was *theoretically* possible that crossing over took place before replication, at the *two-chromosome* stage. This uncertainty was resolved through the genetic analysis of organisms whose four products of meiosis remain together in groups of four called *tetrads*. These organisms, which we met in Chapters 2 and 3, are fungi and unicellular

algae. The products of meiosis of a single tetrad can be isolated, which is equivalent to isolating all four chromatids from a single meiocyte. Tetrad analyses of crosses *in which genes are linked* show many tetrads that contain four different allele combinations. For example, from the cross

$$AB \times ab$$

some (but not all) tetrads contain four genotypes:

$$AB$$

$$Ab$$

$$aB$$

$$ab$$

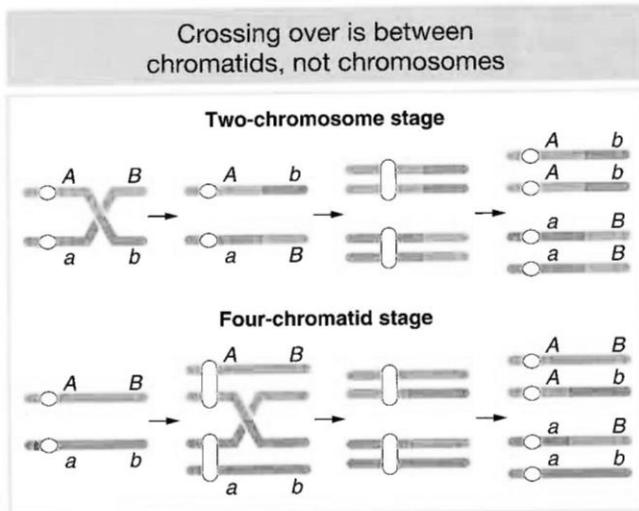


FIGURE 4-5 Crossing over takes place at the four-chromatid stage. Because more than two different products of a single meiosis can be seen in some tetrads, crossing over cannot take place at the two-strand stage (before DNA replication). The white circle designates the position of the centromere. When sister chromatids are visible, the centromere appears unreplicated.

This result can be explained only if crossovers take place at the four-chromatid stage because, if crossovers took place at the two-chromosome stage, there could only ever be a maximum of two different genotypes in an individual tetrad. This reasoning is illustrated in Figure 4-5.

Multiple crossovers can include more than two chromatids

Tetrad analysis can also show two other important features of crossing over. First, in some individual meiocytes, several crossovers can occur along a chromosome pair. Second, in any one meiocyte, these multiple crossovers can exchange material between more than two chromatids. To think about this matter, we need to look at the simplest case: double crossovers. To study double crossovers, we need three linked genes. For example, if the three loci are all linked in a cross such as

$$ABC \times abc$$

many different tetrad types are possible, but some types are informative in the present connection because they can be accounted for only by double crossovers in which more than two chromatids take part. Consider the following tetrad as an example:

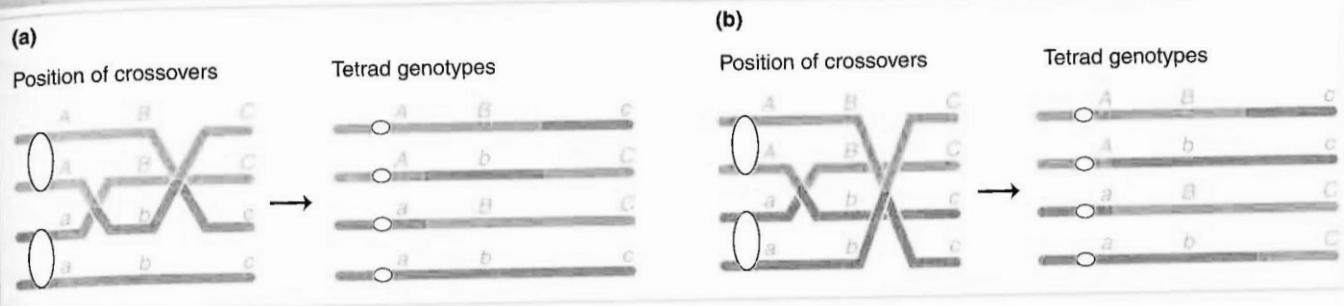
$$\begin{array}{c} ABc \\ AbC \\ aBC \\ abc \end{array}$$

This tetrad must be explained by two crossovers in which *three* chromatids take part, as shown in Figure 4-6a. Furthermore, the following type of tetrad shows that all *four* chromatids can participate in crossing over in the same meiosis (Figure 4-6b):

$$\begin{array}{c} ABc \\ Abc \\ aBC \\ abC \end{array}$$

Therefore, for any pair of homologous chromosomes, two, three, or four chromatids can take part in crossing-over events in a single meiocyte. Note, however, that any single crossover is between two chromatids.

Multiple crossovers can include more than two chromatids



You might be wondering about crossovers between *sister* chromatids. They do occur but are rare. They do not produce new allele combinations and so are not usually considered.

4.2 Mapping by Recombinant Frequency

The frequency of recombinants produced by crossing over is the key to chromosome mapping. Fungal tetrad analysis has shown that, for any two specific linked genes, crossovers take place between them in some, but not all, meiocytes (Figure 4-7). The farther apart the genes are, the more likely that a crossover will take place and the higher the proportion of recombinant products will be. Thus, the proportion of recombinants is a clue to the distance separating two gene loci on a chromosome map.

As stated earlier in regard to Morgan's data, the recombinant frequency was significantly less than 50 percent, specifically 10.7 percent. Figure 4-8 shows the general situation for linkage in which recombinants are less than 50 percent. Recombinant frequencies for different linked genes range from 0 to 50 percent, depending on their closeness. The farther apart genes are, the more closely their recombinant frequencies approach 50 percent, and, in such cases, one cannot decide whether genes are linked or are on different chromosomes. What about recombinant frequencies greater than 50 percent? The answer is that such frequencies are *never* observed, as will be proved later.

Recombinants are produced by crossovers

	Meiotic chromosomes	Meiotic products	
Meioses with no crossover between the genes	A B A B a b a b	○ A B ○ A B ○ a b ○ a b	Parental Parental Parental Parental
	A B A B a b a b	○ A B ○ A b ○ a B ○ a b	Parental Recombinant Recombinant Parental
	A B A B a b a b	○ A B ○ A b ○ a B ○ a b	Parental Recombinant Recombinant Parental
	A B A B a b a b	○ A B ○ A b ○ a B ○ a b	Parental Recombinant Recombinant Parental

FIGURE 4-7
Recombinants arise from meioses in which a crossover takes place between nonsister chromatids.

LaunchPad ANIMATED ART: Meiotic recombination between linked genes by crossing over

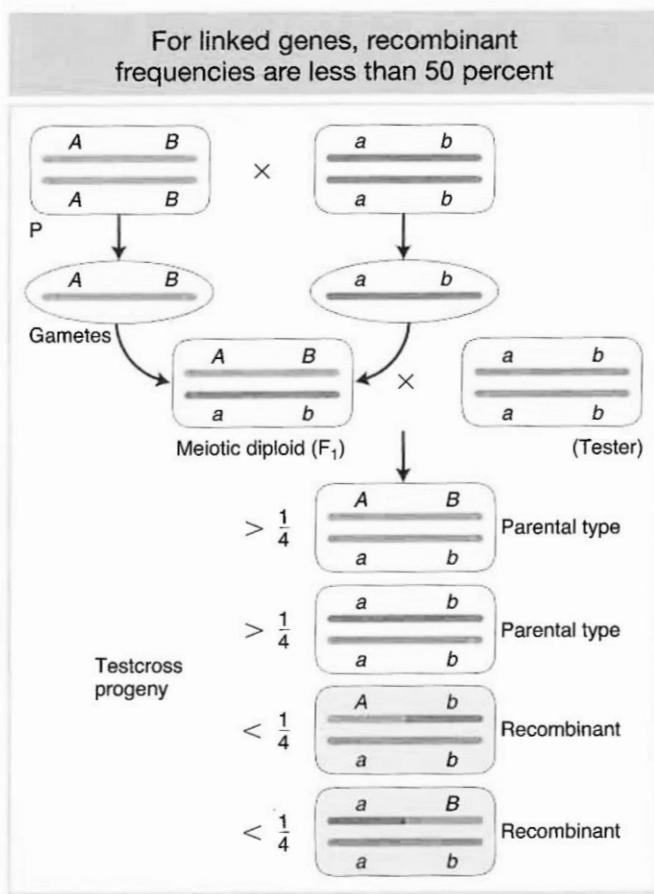


FIGURE 4-8 A testcross reveals that the frequencies of recombinants arising from crossovers between linked genes are less than 50 percent.

Note in Figure 4-7 that a single crossover generates two reciprocal recombinant products, which explains why the reciprocal recombinant classes are generally approximately equal in frequency. The corollary of this point is that the two parental nonrecombinant types also must be equal in frequency, as also observed by Morgan.

Map units

The basic method of mapping genes with the use of recombinant frequencies was devised by a student of Morgan's. As Morgan studied more and more linked genes, he saw that the proportion of recombinant progeny varied considerably, depending on which linked genes were being studied, and he thought that such variation in recombinant frequency might somehow indicate the actual distances separating genes on the chromosomes. Morgan assigned the quantification of this process to an undergraduate student, Alfred Sturtevant, who also became one of the great geneticists. Morgan asked Sturtevant to try to make some sense of the data on crossing over between different linked genes. In one evening, Sturtevant developed a method for mapping genes that is still used today. In Sturtevant's own words, "In the latter part of 1911, in conversation with Morgan, I suddenly realized that the variations in strength of linkage, already attributed by Morgan to differences in the spatial separation of genes, offered the possibility of determining sequences in the linear dimension of a chromosome. I went home and spent most of the night (to the neglect of my undergraduate homework) in producing the first chromosome map."

As an example of Sturtevant's logic, consider Morgan's testcross results with the *pr* and *vg* genes, from which he calculated a recombinant frequency of 10.7 percent. Sturtevant suggested that we can use this percentage of recombinants as a quantitative index of the linear distance between two genes on a genetic map, or **linkage map**, as it is sometimes called.

The basic idea here is quite simple. Imagine two specific genes positioned a certain fixed distance apart. Now imagine random crossing over along the paired homologs. In some meioses, nonsister chromatids cross over by chance in the chromosomal region between these genes; from these meioses, recombinants are produced. In other meiotic divisions, there are no crossovers between these genes; no recombinants result from these meioses. (See Figure 4-7 for a diagrammatic illustration.) Sturtevant postulated a rough proportionality: the greater the distance between the linked genes, the greater the chance of crossovers in the region between the genes and, hence, the greater the proportion of recombinants that would be produced. Thus, by determining the frequency of recombinants, we can obtain a measure of the map distance between the genes. In fact, Sturtevant defined one **genetic map unit (m.u.)** as that distance between genes for which 1 product of meiosis in 100 is recombinant. For example, the **recombinant frequency (RF)** of 10.7 percent obtained by Morgan is defined as 10.7 m.u. A map unit is sometimes referred to as a **centimorgan (cM)** in honor of Thomas Hunt Morgan.

Does this method produce a linear map corresponding to chromosome linearity? Sturtevant predicted that, on a linear map, if 5 map units (5 m.u.) separate genes *A* and *B*, and 3 m.u. separate genes *A* and *C*, then the distance separating *B* and *C* should be either 8 or 2 m.u. (Figure 4-9). Sturtevant found his prediction to be the case. In other words, his analysis strongly suggested that genes are arranged in some linear order, making map distances additive. (There are some minor but not insignificant exceptions, as we will see later.) Since we now know from

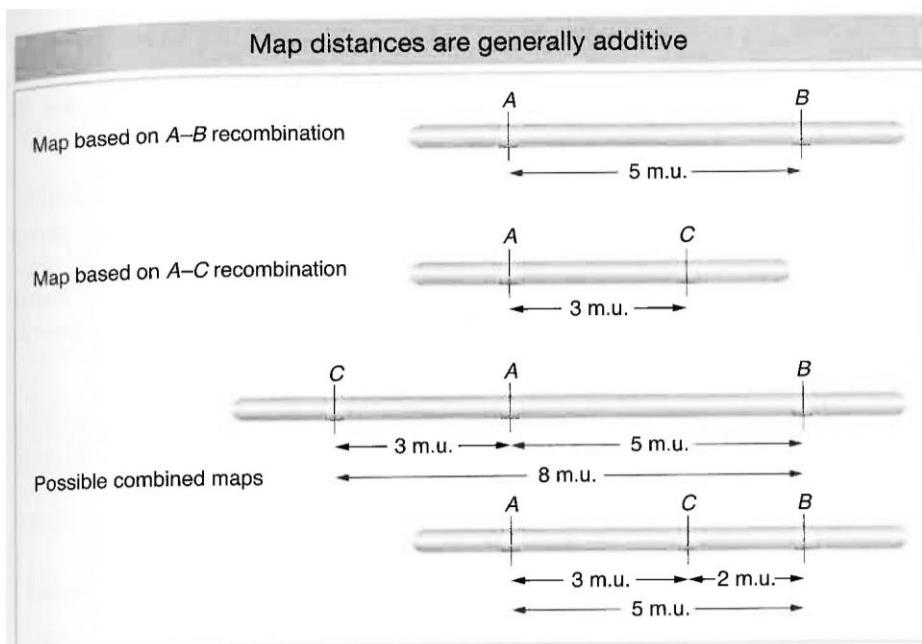
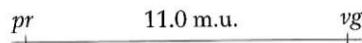


FIGURE 4-9 A chromosome region containing three linked genes. Because map distances are additive, calculation of A–B and A–C distances leaves us with the two possibilities shown for the B–C distance.

molecular analysis that a chromosome is a single DNA molecule with the genes arranged along it, it is no surprise for us today to learn that recombination-based maps are linear because they reflect a linear array of genes.

How is a map represented? As an example, in *Drosophila*, the locus of the eye-color gene and the locus of the wing-length gene are approximately 11 m.u. apart, as mentioned earlier. The relation is usually diagrammed in the following way:



Generally, we refer to the locus of this eye-color gene in shorthand as the "pr locus," after the first discovered mutant allele, but we mean the place on the chromosome where *any* allele of this gene will be found, mutant or wild type.

As stated in Chapters 2 and 3, genetic analysis can be applied in two opposite directions. This principle is applicable to recombinant frequencies. In one direction, recombinant frequencies can be used to make maps. In the other direction, when given an established map with genetic distance in map units, we can predict the frequencies of progeny in different classes. For example, the genetic distance between the *pr* and *vg* loci in *Drosophila* is approximately 11 m.u. So knowing this value, we know that there will be 11 percent recombinants in the progeny from a testcross of a female dihybrid heterozygote in cis conformation (*pr* *vg*/*pr*⁺ *vg*⁺). These recombinants will consist of two reciprocal recombinants of equal frequency: thus, 5.5 percent will be *pr* *vg*⁺ and 5.5 percent will be *pr*⁺ *vg*. We also know that 100 – 11 = 89 percent will be nonrecombinant in two equal classes, 44.5 percent *pr*⁺ *vg*⁺ and 44.5 percent *pr* *vg*. (Note that the tester contribution *pr* *vg* was ignored in writing out these genotypes.)

There is a strong implication that the "distance" on a linkage map is a physical distance along a chromosome, and Morgan and Sturtevant certainly intended to imply just that. But we should realize that the linkage map is a *hypothetical* entity constructed from a purely genetic analysis. The linkage map could have been derived without even knowing that chromosomes existed. Furthermore, at this point in our discussion, we cannot say whether the "genetic distances" calculated by means of recombinant frequencies in any way represent actual physical distances on chromosomes. However, physical mapping has shown that genetic

distances are, in fact, roughly proportional to recombination-based distances. There are exceptions caused by recombination hotspots, places in the genome where crossing over takes place more frequently than usual. The presence of hotspots causes proportional expansion of some regions of the map. Recombination blocks, which have the opposite effect, also are known.

A summary of the way in which recombinants from crossing over are used in mapping is shown in Figure 4-10. Crossovers occur more or less randomly along the chromosome pair. In general, in longer regions, the average number of crossovers is higher and, accordingly, recombinants are more frequently obtained, translating into a longer map distance.

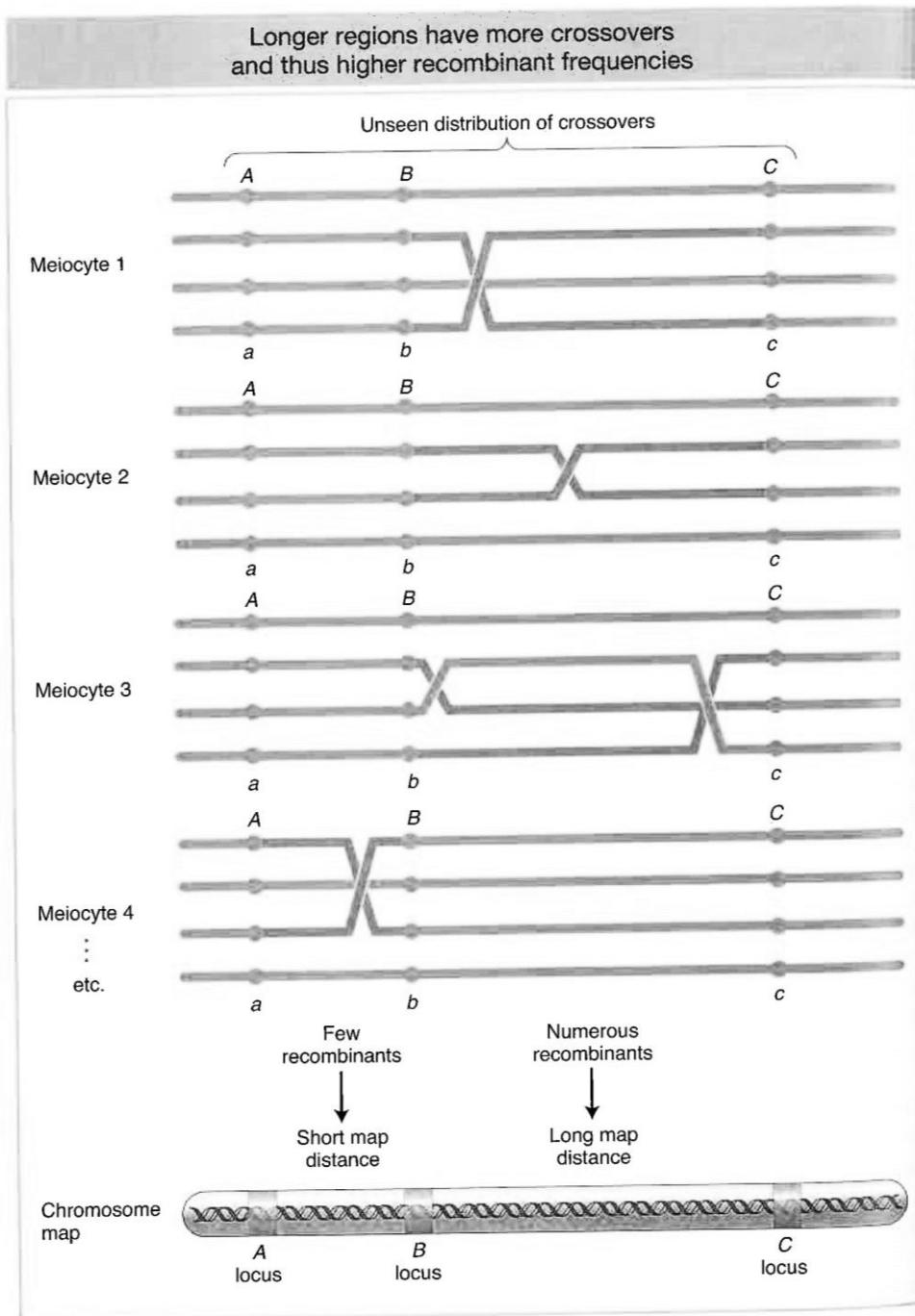


FIGURE 4-10 Crossovers produce recombinant chromatids whose frequency can be used to map genes on a chromosome. Longer regions produce more crossovers. Brown shows recombinants for that interval.

KEY CONCEPT Recombination between linked genes can be used to map their distance apart on a chromosome. The unit of mapping (1 m.u.) is defined as a recombinant frequency of 1 percent.

Three-point testcross

So far, we have looked at linkage in crosses of dihybrids (double heterozygotes) with doubly recessive testers. The next level of complexity is a cross of a trihybrid (triple heterozygote) with a triply recessive tester. This kind of cross, called a **three-point testcross** or a **three-factor cross**, is commonly used in linkage analysis. The goal is to deduce whether the three genes are linked and, if they are, to deduce their order and the map distances between them.

Let's look at an example, also from *Drosophila*. In our example, the mutant alleles are *v* (vermillion eyes), *cv* (crossveinless, or absence of a crossvein on the wing), and *ct* (cut, or snipped, wing edges). The analysis is carried out by performing the following crosses:

P	$v^+/v^+ \cdot cv/cv \cdot ct/ct \times v/v \cdot cv^+/cv^+ \cdot ct^+/ct^+$
	↓
Gametes	$v^+ \cdot cv \cdot t \quad v \cdot cv^+ \cdot ct^+$
F ₁ trihybrid	$v/v^+ \cdot cv/cv^+ \cdot ct/ct^+$

Trihybrid females are testcrossed with triple recessive males:

$v/v^+ \cdot cv/cv^+ \cdot ct/ct^+ \text{♀}$	\times	$v/v \cdot cv/cv \cdot ct/ct^+ \text{♂}$
F ₁ trihybrid female		Tester male

From any trihybrid, only $2 \times 2 \times 2 = 8$ gamete genotypes are possible. They are the genotypes seen in the testcross progeny. The following chart shows the number of each of the eight gametic genotypes in a sample of 1448 progeny flies. The columns alongside show which genotypes are recombinant (R) for the loci taken two at a time. We must be careful in our classification of parental and recombinant types. Note that the parental input genotypes for the triple heterozygotes are $v^+ \cdot cv \cdot ct$ and $v \cdot cv^+ \cdot ct^+$; any combination other than these two constitutes a recombinant.

Gametes	Recombinant for loci		
	<i>v</i> and <i>cv</i>	<i>v</i> and <i>ct</i>	<i>cv</i> and <i>ct</i>
$v \cdot cv^+ \cdot ct^+$	580		
$v^+ \cdot cv \cdot ct$	592		
$v \cdot cv \cdot ct^+$	45	R	
$v^+ \cdot cv^+ \cdot ct$	40	R	
$v \cdot cv \cdot ct$	89	R	R
$v^+ \cdot cv^+ \cdot ct^+$	94	R	R
$v \cdot cv^+ \cdot ct$	3		R
$v^+ \cdot cv \cdot ct^+$	5		R
	1448	268	191
			93

Let's analyze the loci two at a time, starting with the *v* and *cv* loci. In other words, we look at just the first two columns under "Gametes" and cover up the third one. Because the parents for this pair of loci are $v^+ \cdot cv$ and $v \cdot cv^+$, we know that the recombinants are by definition $v \cdot cv$ and $v^+ \cdot cv^+$. There are $45 + 40 + 89 + 94 = 268$

of these recombinants. Of a total of 1448 flies, this number gives an RF of 18.5 percent.

For the *v* and *ct* loci, the recombinants are *v*·*ct* and *v*⁺·*ct*⁺. There are $89 + 94 + 3 + 5 = 191$ of these recombinants among 1448 flies, and so the RF = 13.2 percent.

For *ct* and *cv*, the recombinants are *cv*·*ct*⁺ and *cv*⁺·*ct*. There are $45 + 40 + 3 + 5 = 93$ of these recombinants among the 1448, and so the RF = 6.4 percent.

Clearly, all the loci are linked, because the RF values are all considerably less than 50 percent. Because the *v* and *cv* loci have the largest RF value, they must be farthest apart; therefore, the *ct* locus must lie between them. A map can be drawn as follows:



The testcross can be rewritten as follows, now that we know the linkage arrangement:

$$v^+ ct\ cv/v\ ct^+ cv^+ \times v\ ct\ cv/v\ ct\ cv$$

Note several important points here. First, we have deduced a gene order that is different from that used in our list of the progeny genotypes. Because the point of the exercise was to determine the linkage relation of these genes, the original listing was of necessity arbitrary; the order was simply not known before the data were analyzed. Henceforth, the genes must be written in correct order.

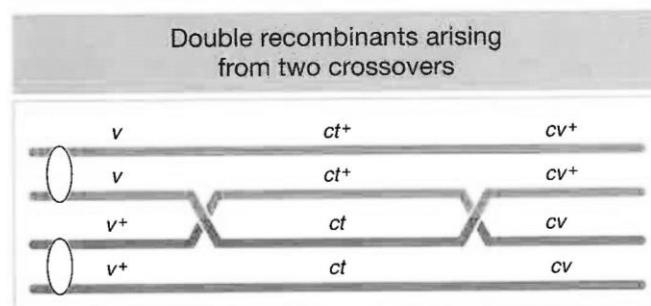
Second, we have definitely established that *ct* is between *v* and *cv*. In the diagram, we have arbitrarily placed *v* to the left and *cv* to the right, but the map could equally well be drawn with the placement of these loci inverted.

Third, note that linkage maps merely map the loci in relation to one another, with the use of standard map units. We do not know where the loci are on a chromosome—or even which specific chromosome they are on. In subsequent analyses, as more loci are mapped in relation to these three, the complete chromosome map would become “fleshed out.”

KEY CONCEPT Three-point (and higher) testcrosses enable geneticists to evaluate linkage between three (or more) genes and to determine gene order, all in one cross.

A final point to note is that the two smaller map distances, 13.2 m.u. and 6.4 m.u., add up to 19.6 m.u., which is greater than 18.5 m.u., the distance calculated for *v* and *cv*. Why? The answer to this question lies in the way in which we have treated the two rarest classes of progeny (totaling 8) with respect to the recombination of *v* and *cv*. Now that we have the map, we can see that these two rare classes are in fact double recombinants, arising from two crossovers (Figure 4-11). However, when we calculated the RF value for *v* and *cv*, we did not count the *v*·*ct*⁺ and *v*⁺·*ct*⁺·*cv* genotypes; after all, with regard to *v* and *cv*, they are parental combinations (*v*·*cv*⁺ and *v*⁺·*cv*). In light of our map, however, we see that this oversight led us to underestimate the distance between the *v* and the *cv* loci. Not only should we have counted the two rarest classes, we should have counted each of them *twice* because each represents double recombinants. Hence, we can correct the value by adding the numbers $45 + 40 + 89 + 94 + 3 + 3 + 5 + 5 = 284$. Of the total of 1448, this number is exactly 19.6 percent, which is identical with the sum of the two component values. (In practice, we do not need to do this calculation, because the sum of the two shorter distances gives us the best estimate of the overall distance.)

FIGURE 4-11 Example of a double crossover between two chromatids. Notice that a double crossover produces double-recombinant chromatids that have the parental allele combinations at the outer loci. The position of the centromere cannot be determined from the data. It has been added for completeness.



Deducing gene order by inspection

Now that we have had some experience with the three-point testcross, we can look back at the progeny listing and see that, for trihybrids of linked genes, *gene order* can usually be deduced by inspection, without a recombinant frequency analysis. Typically, for linked genes, we have the eight genotypes at the following frequencies:

- two at high frequency
- two at intermediate frequency
- two at a different intermediate frequency
- two rare

Only three gene orders are possible, each with a different gene in the middle position. It is generally true that the double-recombinant classes are the smallest ones, as listed last here. Only one order is compatible with the smallest classes' having been formed by double crossovers, as shown in Figure 4-12; that is, only one order gives double recombinants of genotype $v\ ct\ cv^+$ and $v^+ \ ct^+ \ cv$. A simple rule of thumb for deducing the gene in the middle is that it is the allele pair that has "flipped" position in the double-recombinant classes.

Interference

Knowing the existence of double crossovers permits us to ask questions about their possible interdependence. We can ask, Are the crossovers in adjacent chromosome regions independent events or does a crossover in one region affect the likelihood of there being a crossover in an adjacent region? The answer is that, generally, crossovers inhibit each other somewhat in an interaction called **interference**. Double-recombinant classes can be used to deduce the extent of this interference.

Interference can be measured in the following way. If the crossovers in the two regions are independent, we can use the product rule (see page 94) to predict the frequency of double recombinants: that frequency would equal the product of the recombinant frequencies in the adjacent regions. In the $v\text{-}ct\text{-}cv$ recombination data, the $v\text{-}ct$ RF value is 0.132 and the $ct\text{-}cv$ value is 0.064; so, if there is no interference, double recombinants might be expected at the frequency $0.132 \times 0.064 = 0.0084$ (0.84 percent). In the sample of 1448 flies, $0.0084 \times 1448 = 12$ double recombinants

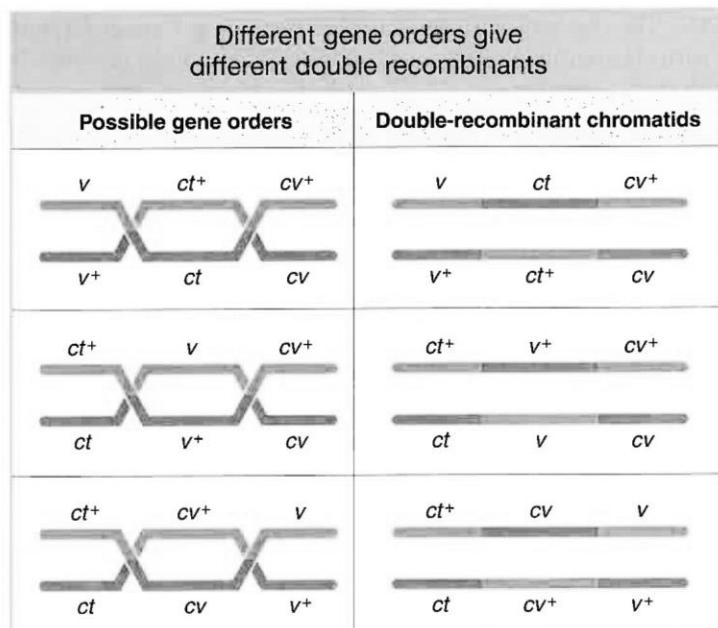


FIGURE 4-12 The three possible gene orders shown on the left yield the six products of a double crossover shown on the right. Only the first possibility is compatible with the data in the text. Note that only the nonsister chromatids taking part in the double crossover are shown.

are expected. But the data show that only 8 were actually observed. If this deficiency of double recombinants were consistently observed, it would show us that the two regions are not independent and suggest that the distribution of crossovers favors singles at the expense of doubles. In other words, there is some kind of interference: a crossover does reduce the probability of a crossover in an adjacent region.

Interference is quantified by first calculating a term called the **coefficient of coincidence (c.o.c.)**, which is the ratio of observed to expected double recombinants. Interference (I) is defined as $1 - \text{c.o.c.}$. Hence,

$$I = 1 - \frac{\text{observed frequency, or}}{\text{number, of double recombinants}} \\ \frac{\text{expected frequency, or}}{\text{number, of double recombinants}}$$

In our example,

$$I = 1 - \frac{8}{12} = \frac{4}{12} = \frac{1}{3}, \text{ or } 33 \text{ percent}$$

In some regions, there are never any observed double recombinants. In these cases, $\text{c.o.c.} = 0$, and so $I = 1$ and interference is complete. Interference values anywhere between 0 and 1 are found in different regions and in different organisms.

You may have wondered why we always use heterozygous females for test-crosses in *Drosophila*. The explanation lies in an unusual feature of *Drosophila* males. When, for example, *pr vg/pr⁺ vg⁺* males are crossed with *pr vg/pr vg* females, only *pr vg/pr⁺ vg⁺* and *pr vg/pr vg* progeny are recovered. This result shows that there is no crossing over in *Drosophila* males. However, this absence of crossing over in one sex is limited to certain species; it is not the case for males of all species (or for the heterogametic sex). In other organisms, there is crossing over in XY males and in WZ females. The reason for the absence of crossing over in *Drosophila* males is that they have an unusual prophase I, with no synaptonemal complexes. Incidentally, there is a recombination difference between human sexes as well. Women show higher recombinant frequencies for the same autosomal loci than do men.

With the use of a reiteration of the preceding recombination-based techniques, maps have been produced of thousands of genes for which variant (mutant) phenotypes have been identified. A simple illustrative example from the tomato is shown in Figure 4-13. The tomato chromosomes are shown in Figure 4-13a, their numbering in Figure 4-13b, and recombination-based gene maps in Figure 4-13c. The chromosomes are shown as they appear under the microscope, together with chromosome maps based on linkage analysis of various allelic pairs shown with their phenotypes.

1.4 Lecturer's notes: Quantitative Genetics



Brief Introduction to Plant and Animal Breeding

Peter Sørensen & Guillaume Ramstein

2023-01-12

The Quantitative Genetics lecture notes consist of the following chapters:

Chapter 1: Brief Introduction to Plant and Animal Breeding
Chapter 2: Basic concepts in Quantitative Genetics used in Breeding
Chapter 3: Estimation of Genetic Parameters used in Breeding
Chapter 4: Estimation of Breeding Values
Chapter 5: Estimation of Genomic Breeding Values

Appendix 1: Using R for Statistical Analyses of Quantitative Traits.

The material provided in these notes is in part from the following sources (CC-BY-SA 4.0, <https://creativecommons.org/licenses/by-sa/4.0/>): https://en.wikipedia.org/wiki>Selective_breeding
https://charlotte-ngs.github.io/lbgfs2021/cn/01_intro_lbg.pdf

1 Brief Introduction to Breeding in Animals and Plants

Selective breeding (also called artificial selection) is the process by which humans choose which parents will sexually reproduce and make offspring. Domesticated animals are known as breeds, normally created by a professional breeder. Domesticated plants are known as varieties or cultivars. Two purebred animals of different breeds produce a crossbreed, and crossbred plants are called hybrids. Flowers, vegetables and fruit trees may be bred by amateurs or professional breeders: major crops are usually produced by professional breeders, in academia or industry.

In both animal and plant breeding, crossing techniques such as inbreeding, linebreeding, and outcrossing are utilized. Charles Darwin (https://en.wikipedia.org/wiki/Charles_Darwin) discussed how selective breeding had been successful in producing change over time in his 1859 book, On the Origin of Species. Its first chapter discusses selective breeding and domestication of animals like pigeons, cats, cattle, and dogs. Darwin used artificial selection as a springboard to introduce and support the theory of natural selection.

The deliberate exploitation of selective breeding to produce desired results has become very common in agriculture and experimental biology.

Selective breeding can be unintentional. In plants, it may result from the process of human cultivation, and may produce unintended – desirable or undesirable – results. For example, in some grains, an increase in seed size may have resulted from certain ploughing practices rather than from the intentional selection of larger seeds. Most likely, there has been an interdependence between natural and artificial factors that have resulted in plant domestication.

1.1 Fundamental Questions in Selective Breeding

Natural selection and selective breeding can both cause changes in animals and plants. The difference between the two is that natural selection happens naturally, but selective breeding only occurs when humans intervene. For this reason, selective breeding is sometimes called artificial selection. Selective breeding takes place over many generations. These are the main steps involved:

- Which traits should we breed for?
- How do we select the *best* breeding individuals?
- What can breeders do to obtain the *best* breeding individuals?

The term *best* is relative, because there is no *best* breeding for all situations and all environments. Breeding individuals that show high performance in one environment, may not perform as well in a different environment. For example, Holstein cows in Europe or North America, are able to produce a lot of milk, but they have difficulties surviving in Africa. Knowing that the environment plays an important role for animals and plants, we usually assume that breeding populations are more or less adapted to their environment.

Breeding individuals are usually described or characterized in terms of appearance and/or performance. In any case, we will be talking about **traits** where any trait is a *categorical* or *continuous* characteristic of an breeding individual. Examples of traits are:

Type of traits	Animals	Plants
Categorical	Coat color Size Muscling Leg set Udder conformation	Grain/flower/leaf color Branching pattern Disease
Continuous	Body weight Milk yield Protein and fat yield Carcass weight	Plant height Grain yield Biomass yield Nutritional content

Note, it is important to distinguish between the categorical or continuous *values* of a trait which might be, e.g., red coat color or 343 kg of body weight, and the traits themselves which are just coat color or body weight. The measured values of a trait are also called **phenotypes**.

1.2 Genotypes and Phenotypes

In selective breeding we are mainly concerned with changing the populations at the genetic level. The reason why we are interested in changing a population genetically is because parents do not pass their phenotypes per se to their offspring. Instead, they pass on a random sample of their genes to their offspring. For each offspring every parent transmits a different sample of their genes. From a genetic point of view, we want to know not only the most desirable phenotype, but also the most desirable genotypes. From the central dogma of molecular biology (https://en.wikipedia.org/wiki/Central_dogma_of_molecular_biology), it follows that an animal's genotype provides the genetic basis for phenotypes. The relationship between phenotypes (P) and genotypes (G) can be summarized by the following equation:

$$P = G + E \tag{1}$$

where E represents the **environmental effects**. These may include external effects like climate and soil composition (abiotic factors) and exposure to beneficial or pathogenic microbes (biotic factors). They may also include developmental effects like maternal effects, or simply measurement errors (random fluctuations in measurement of phenotypes). Because we want to change our populations at the genetic level, we are interested in the effect of genotypes (G). In most cases, we are not able to directly observe or measure G . But we will see later on how we can estimate G based on measurements and observations of P and based on estimates of E . In particular, we will see how **breeding values**, the part of G that is transmitted over generation, are estimated and used by breeders to perform crosses and improve populations. Those tools are being described in the following section.

1.3 Improvement of Breeding Populations

The purpose of selective breeding is to improve the traits selected for, in a breeding population. Once a breeding individual is conceived, its genotype is fixed and cannot be improved anymore. Breeders can improve populations at the genetic level through the following steps:

- **evaluation** of individuals in the current generation
- **selection** of the ‘best’ individuals (for phenotypes of interest)
- **mating** of selected individuals to generate the next (improved) generation

1.3.1 Evaluation

Breeding value estimation is a fundamental component of breeding programmes, in which the estimated breeding value of each individual informs subsequent selection decisions. The breeding value for an individual is the effect its genes will have in the population. Estimated breeding values are used for 1) comparing individuals in the breeding population and selecting parents for the next generation, 2) predicting the consequences of selection decisions, and 3) describing genetic differences over time (result of previous selection).

1.3.2 Selection

Selection is the process of choosing individuals in the current generation, as parents for the next generation. The application of selection in a certain population over a certain time changes the breeding individuals in that population at the genetic level. The most familiar form of selection is natural selection which occurs in natural and wildlife populations. Natural selection is one of the great forces of evolution and it also affects domestic animal and plant populations. All animals or plants with lethal genetic defects are naturally selected against, i.e., they do not survive before the breeder gets to evaluate them.

Although natural selection cannot be ignored, the focus of animal and plant breeders is **artificial selection**. The idea behind artificial selection is simple. For a given trait all individuals in a breeding population are ranked according to their estimated breeding value. From this list, the top-ranking breeding individuals are used as parents for the next generation. In most breeding populations, breeders are interested in measuring more than one trait. When considering more than one trait, the question is how to come up with the ranking for the individuals to be selected as potential parents. There are several strategies to produce such a ranking based on a number of traits. It has been shown that using a weighted mean of the breeding values of all traits, which is called an **aggregate genotype**, to rank all animals is an optimal procedure to be used as selection criterion.

1.3.3 Mating

The last step of each breeding cycle is **mating**. We must decide which of the selected males and females are mated with each other. There are a number of different rules that can be followed. The application of a given set of rules is summarized as a mating system. There are three reasons for using a specific mating system.

1. producing offspring with extreme breeding values. When parents with extreme breeding values (high or low) are mated, offspring with extreme phenotypes can be expected. This is mostly used when a given trait is to be changed in one direction
2. making use of complementarity in parental traits. When neither of the parents is optimal, a mix of traits can be desirable. Crossing parents of different breeds is called **crossbreeding**. Hybrid vigor, or *heterosis*, may then occur, when offspring performance exceeds the performance of the parents.

There might also be other aspects that influence a mating system, e.g. to restrict the level of inbreeding or to consider optimum genetic contribution theory.

Introduction to Basic Concepts in Quantitative Genetics

Guillaume Ramstein & Peter Sørensen

2023-01-12

Contents

Learning objective:	1
1 Infinitesimal model	2
2 Genetic effects at each locus	2
2.1 Additive effect	3
2.2 Dominance effect	3
2.3 Genotypic value versus breeding value	3
3 Distribution of genotypic and phenotypic values	4
3.1 Single-locus model	4
3.2 Multiple-locus model	4
4 Genetic parameters	4
4.1 Genetic variance	4
4.2 Heritability	6
4.3 Genetic correlation	7
5 Basic questions remain	7

Learning objective:

This section introduces the basic concepts in quantitative genetics such as:

- Genetic value and variance for a quantitative trait
- Genetic parameters (heritability, genetic variance and correlation)
- Difference between genotypic values and breeding values
- Infinitesimal model

These concepts are relevant for a range of genetic and statistical analyses of complex traits and diseases in animal and plant populations, including:

- Estimating the heritability of a trait (the part of its variability due to genetics)
- Estimating breeding values by pedigree or genomic information
- Selection of breeding individuals based on estimated breeding values
- Prediction of selection response based on estimated heritability (breeder's equation)

Quantitative genetics is the study of **complex traits**, influenced by many genetic loci. Quantitative genetics is based on models in which many genes influence the trait, and in which non-genetic factors may also be important. Quantitative traits such as size, obesity or longevity vary greatly among individuals. Their phenotypes are continuously distributed phenotypes and do not show simple Mendelian inheritance (i.e., their phenotypes are not distributed in discrete categories determined by one or a few genes). The quantitative genetics framework can also be used to analyze discrete traits like litter size (which consists of discrete counts

like 0, 1, 2, 3, ...) or binary traits like survival to adulthood (which consists of 0 or 1, ‘dead’ or ‘alive’, etc.), provided that they have a polygenic basis (i.e., they are determined by many genes). The quantitative genetics approach has diverse applications: it is fundamental to an understanding of variation and covariation among relatives in natural and managed populations; it is also used as basis for selective breeding methods in animal and plant populations (<https://doi.org/10.1098/rstb.2009.0203>).

1 Infinitesimal model

The **infinitesimal model** is a widely used genetic model in quantitative genetics. Originally developed in 1918 by Ronald Fisher, it is based on the idea that variation in a quantitative trait is influenced by an infinitely large number of genes, each of which makes an infinitely small (infinitesimal) contribution to the phenotype, as well as by environmental (non-genetic) factors. In the most basic model the phenotype (P) is the sum of genetic effects (G), and environmental effects (E):

$$P = G + E \quad (1)$$

The genotypic value (G) in the model can be split into additive genetic value (A), dominance value (D), and epistatic value (I) such that the expanded infinitesimal model becomes:

$$P = A + D + I + E \quad (2)$$

The additive genetic value (A) is the **breeding value**, the part of the genotypic value that is passed along to the next generation. The breeding value is the sum of individual allele effects. The dominance value (D) is caused by interactions between alleles *within loci*. The epistatic value (I) is caused by interactions *between loci*. Even though epistasis is relevant in biology, it is often ignored in quantitative genetics models, for simplicity of the models.

The genotypic effect may also depend on the environment in which they are expressed (e.g., in plants a drought-tolerance gene may have a favorable effect on grain yield under water-limited conditions, but may be useless under irrigation). Therefore we may consider an extended version of the infinitesimal model where the phenotype (P) consists of the genotypic value (G), environmental value (E), and genotype-environment interactions ($G \times E$):

$$P = G + E + G \times E \quad (3)$$

In practice, genotype-environment interactions may be important for the phenotype of individuals (especially in plant breeding), but for the sake of simplicity we will ignore them in the remainder of this section.

Hereafter, by ignoring epistasis and genotype-environment interactions (for simplicity), we will assume the following model:

$$P = A + D + E \quad (4)$$

2 Genetic effects at each locus

Genetic effects are the contributions of loci to the genotypic value (G). Additive effects of loci contribute to the breeding value (A), and dominance effects of loci contribute to the dominance value (D). In absence of epistasis, A is the sum of additive effects over loci; similarly, D is the sum of dominance effects over loci.

In this course, we will always consider individuals that are diploid, like most animals and many plant species (e.g., maize, soybean, barley). So, each locus exists under two copies. How do the individual's copies combine into a genotype?

Consider two possible alleles A_1 and A_2 for each copy at the locus, with respective effects +1 and -1. If genetic effects are entirely additive, then the value of each possible genotype is the sum of the effects of each copy, i.e., the value of genotypes at the locus are -2 (if the individual is A_2A_2), 0 (if it is A_2A_1 or A_1A_2), or +2 (if it is A_1A_1).

However, the value of each genotype will generally depend on the combination of alleles within one locus ($G = A + D$) or even across multiple loci (in presence of epistasis). For example, in presence of dominance, the value of each possible genotype may be -2 if the individual is A_2A_2 , +1 if it is A_2A_1 (or A_1A_2), and +2 if it is A_1A_1 .

2.1 Additive effect

The additive effect of a locus is the sum of its allele effects. Quite confusingly, allele effects depend on the population, because they depend on the frequency of genotypes in the population!

For example, assume that genotypes at one locus have values -2 (A_2A_2), +1 (A_2A_1) and +2 (A_1A_1). In a population consisting of 25% A_2A_2 , 50% A_2A_1 and 25% A_1A_1 , you would expect the A_1 allele in a A_1A_2 genotype 2/3 of the time, and you would expect the A_1 allele in a A_1A_1 genotype 1/3 of the time. In another population consisting of 90% A_2A_2 , 18% A_2A_1 and 2% A_1A_1 , you would expect the A_1 allele in a A_1A_2 genotype about 95% of the time, and in a A_1A_1 genotype only about 5% of the time. Therefore, the effect of the A_1 allele, averaged over genotypes in which it is found (A_1A_2 or A_1A_1), will not be the same from one population to another. The concept of allele effects is fundamental to quantitative genetics. However, it is one of the most confusing, precisely because of this dependence on genotype frequencies.

2.2 Dominance effect

Dominance genetic effects are the interactions among alleles at a given locus. This is an effect that is extra to the sum of the additive allele effects. Each genotype has its own dominance effect, denoted by δ_{ij} , for the specific combination of alleles i and j, (e.g., $\delta_{A_1A_2}$), and each of them are non-zero quantities.

2.3 Genotypic value versus breeding value

For selective breeding purposes additive genetic effects are of primary interest. This is because additive effects generally make most of the genotypic effects, and the allelic effects are passed directly to offspring while the other genetic effects are not transmitted to the progeny. Moreover, non-additive genetics effects (like dominance effects) are generally smaller in magnitude. The sum of additive effects across loci on a quantitative trait is the breeding value.

- Breeding value = the value of genes to progeny (additive allele effects only)
- Genotypic value = the value of genes to self (which includes additive, dominance and possibly epistatic effects)

The difference between genotypic value and breeding value is largely due to the dominance value (the sum of dominance effects across loci). However, an individual cannot pass its dominance value to the next generation as it only transmits one allele (e.g., an A_1A_2 heterozygote will either transmit an A_1 gamete or an A_2 gamete to one of its progeny, but not both!). When offspring are generated by crossing (or random mating), each parent contributes a single allele at each locus to its offspring, and hence only passes along a part of its genotypic value. This part is determined by the allele effects (which make up the breeding value). However, any favorable interaction between alleles is not passed along to their offspring.

3 Distribution of genotypic and phenotypic values

Quantitative traits do not behave according to simple Mendelian inheritance laws. More specifically, their inheritance cannot be explained by the genetic segregation of one or a few genes. Even though Mendelian inheritance laws accurately depict the segregation of genotypes in a population, they are not tractable with the large number of genes which typically affect quantitative traits.

To better understand the infinitesimal model, assume Mendelian inheritance to occur at every locus in the genome. Let's say there are 30,000 gene loci in the genome. If we assume that there are only two alleles (3 possible genotypes) per locus, and gene loci segregate independently, then the number of possible genotypes (considering all loci simultaneously) would be 3^{30000} which is large enough to give the "illusion" of an infinite number of loci. Furthermore each of these loci contribute additive and dominance effects, in addition to epistatic effects (which we neglect in our models, for the sake of simplicity).

3.1 Single-locus model

First we may consider how to model the genetic basis of a quantitative trait when only one locus affects the trait of interest. We call this the single-locus model. The distribution of genotypic values in the population is then discrete, and depends on genotype frequencies, which in turn depend on allele frequencies of A_1 and A_2 .

In addition to the genotypic value (G), the phenotypic value (P) is influenced by the environmental value (E). Generally, E is assumed to be normally distributed. Therefore, in a single-locus model, P is normally distributed around each of the possible value of G.

3.2 Multiple-locus model

Now we will consider a multiple-locus model. When several loci are causal (i.e., they have an effect on a certain trait), then we talk about a **polygenic model**. Letting the number of causal loci tend to infinity, the resulting model is the infinitesimal model. From a statistical point of view, the breeding values in an infinitesimal model are considered random with a known distribution. Due to the central limit theorem, this distribution tends to a normal distribution, because of the *approximately* infinite number of causal loci. The central limit theorem says that the distribution of any sum of many small random effects is approximately normal. In our case, where a given trait of interest is thought to be influenced by a large number of genetic loci, each having a small effect, the sum of the additive effects over all loci is approximately normally distributed, i.e., the breeding value (A) is approximately normally distributed.

The histograms below show a better approximation to the normal distribution for breeding values (summed allele effects at causal loci), as the number of causal loci increases. In practice, 100 independently segregating causal loci may be large enough, so that the infinitesimal model (and the normal approximation in genomic models) is accurate enough for predictions.

4 Genetic parameters

Fisher (1918) and Wright (1921) have introduced fundamental statistical methods in quantitative genetics:

- analysis of variance: the partition of phenotypic variation into heritable (A) and non-heritable components (D, and E).
- resemblance among relatives: the estimations of the proportion of loci shared by relatives under the infinitesimal model.

4.1 Genetic variance

In the model proposed by Fisher (1918), Cockerham (1954) and Kempthorne (1954), covariance among relatives is described by the additive genetic variance V_A (variance of additive genetic effects, or breeding values) and dominance variance V_D (variance of interaction effects between alleles in the same locus) (Falconer

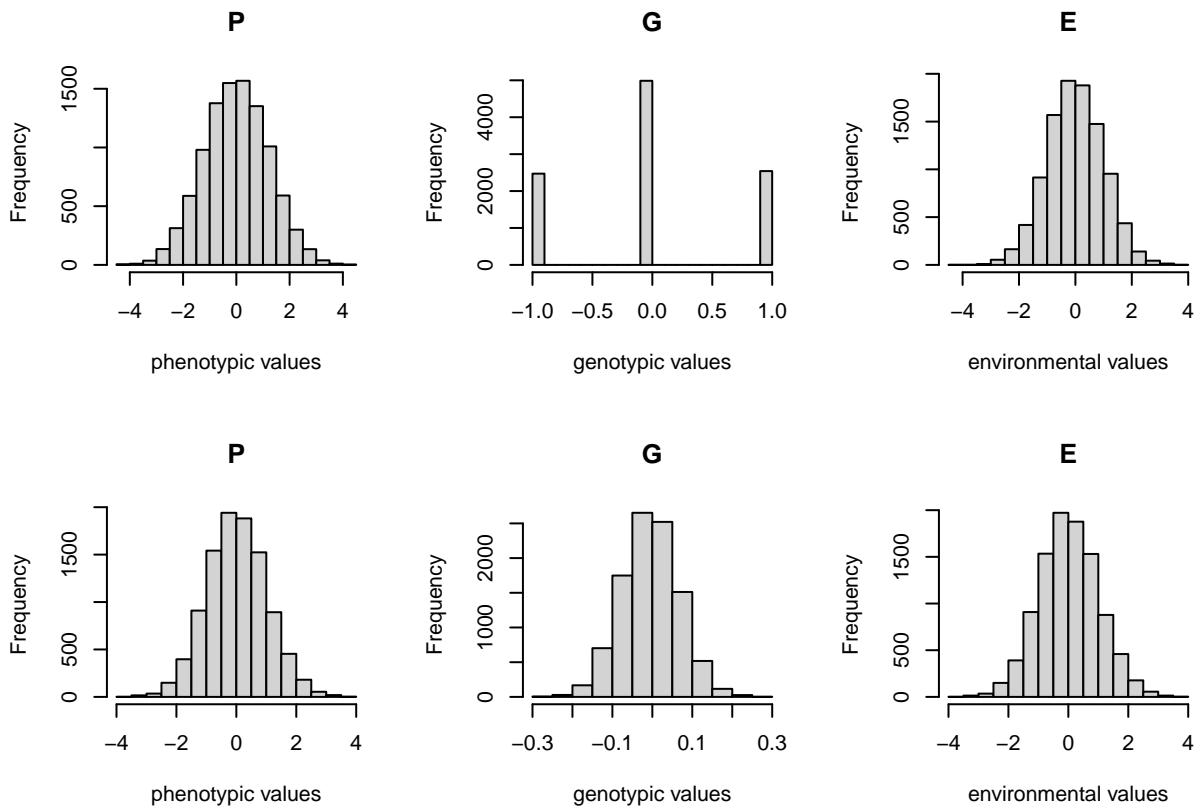


Figure 1: Distribution of genotypic and phenotypic values for a quantitative trait influenced by a single locus model (top panel) or multiple loci (bottom panel)

& Mackay 1996; Lynch & Walsh 1998). Importantly, these partitions rely on the assumption of Hardy-Weinberg equilibrium: random mating, large population size (i.e. no inbred individuals), linkage equilibrium (independent segregation of loci, which requires many generations to achieve for tightly linked genes), and no selection.

$$\begin{aligned} V_P &= V_G + V_E \\ &= V_A + V_D + V_E \end{aligned} \tag{5}$$

$$\begin{aligned} \sigma_P^2 &= \sigma_G^2 + \sigma_E^2 \\ &= \sigma_A^2 + \sigma_D^2 + \sigma_E^2 \end{aligned} \tag{6}$$

Many more terms may be included, such as epistasis, genotype \times environment interactions, and maternal genetic effects (the effect of the mother on its offspring's phenotype). The model has unlimited opportunities for complexity. This is a strength, in that it is all-accommodating, and a weakness, in that datasets may allow to partition only a few components. In practice, assumptions must be made to reduce the complexity of the resemblance among relatives. Usually, the resemblance among relatives is assumed to depend only on additive genetic variance V_A and dominance variance V_D , so that the following sources of covariation are ignored:

- Epistatic variance (interaction effects among loci are small compared to additive and dominance effects)
- Environmental variance (effects of **shared environments** are assumed to be small enough)
- etc.

4.2 Heritability

The models and summary statistics defined by Fisher and Wright have remained at the heart of quantitative genetics, not least because they provide ways to make predictions of important quantities, such as

- Breeding value (A), the expected performance of an individual's offspring
- Broad-sense heritability, the ratio of total genetic variance V_G to the overall phenotypic variance V_P :

$$\begin{aligned} H^2 &= V_G/V_P \\ &= (V_A + V_D)/V_P \\ H^2 &= \sigma_G^2/\sigma_P^2 \\ &= (\sigma_A^2 + \sigma_D^2)/\sigma_P^2 \end{aligned}$$

- Narrow-sense heritability, the ratio of additive genetic variance V_A to the overall phenotypic variance V_P :

$$\begin{aligned} h^2 &= V_A/V_P \\ h^2 &= \sigma_A^2/\sigma_P^2 \end{aligned} \tag{7}$$

- The response to artificial or natural selection, the increase (or decrease) in the genotypic value of individuals due to selection, from one generation to the next

4.3 Genetic correlation

In a general quantitative genetic model, in which, for each individual, two traits (P_1 and P_2) are each defined as the sum of a genotypic value (G_1 and G_2) and an environmental value (E_1 and E_2):

$$P_1 = G_1 + E_1 \quad (8)$$

$$P_2 = G_2 + E_2 \quad (9)$$

The genotypic correlation ($\rho_{G_{12}}$) between the traits is defined as:

$$\rho_{G_{12}} = \frac{\sigma_{G_{12}}}{\sqrt{\sigma_{G_1}^2 \sigma_{G_2}^2}}$$

where $\sigma_{G_{12}}$ is the genotypic covariance between traits, $\sigma_{G_1}^2$ and $\sigma_{G_2}^2$ are the genotypic variances of the two traits in the population. The genotypic correlation conveys the resemblance between genetic effects across traits. Such resemblance may be due to pleiotropy (the same gene impacts the two traits simultaneously) or linkage disequilibrium (different genes impacts different traits, but they are co-inherited). The genotypic correlation is an important consideration when breeding for two or more traits, whose genotypic effects may be positively correlated (e.g., body size and height) or negatively correlated (e.g., grain yield and resilience).

5 Basic questions remain

In view of the assumed complexity of the underlying gene actions, involving many loci with unknown effects and interactions, much quantitative genetic analysis has, unashamedly, been at a level of the ‘black box’.

On the premise that many genes and environmental factors interact to impact the trait, it will be difficult to determine the action of individual causal genes. Many basic questions remain:

- What do genes do? What is the underlying biology of these allele effects we talked about?
- How do genes interact? What about those complex biological pathways, whose complexity is overlooked here?
- On what traits does natural selection act? And why is there still genotypic variation, in traits under (natural or artificial) selection?
- Can we expect continued genetic improvement in selection programmes?

Ultimately, we want to know which genes impact the phenotypes, understand the biology of genetic effects (structural or regulatory), and even understand how specific mutations (nucleotide substitution, deletions, copy number variant, etc.) result in allele effects.

References

Estimation of Genetic Parameters

Guillaume Ramstein & Peter Sørensen

2023-01-15

Contents

Learning objective:	1
1 Introduction	1
1.1 Genetic models and variance components	2
1.2 Genetic parameters	2
1.3 Data required for estimating genetic parameters	3
2 Methods for estimation of genetic parameters	3
2.1 Parent-offspring regression on pairs of observations	3
2.2 ANOVA on family data	4
2.3 Restricted Maximum Likelihood on general pedigree	4
3 When to estimate variance components?	8

Learning objective:

This section introduces the basic concepts for estimating genetic parameters:

- genetic model used to define genetic parameters
- different methods, data sources and experimental designs for estimating genetic parameters
- use of genetic relationships for estimating genetic parameters
- importance of genetic parameter estimates in breeding
- knowing when estimation of genetic parameters is required

1 Introduction

Genetic parameters describe the variability of genetic effects: how much genetic effects vary, relative to non-genetic effects (genetic variances and heritability); how correlated genetic effects are, between traits (genetic correlation). The estimation of genetic parameters is a central issue in animal and plant breeding. Genetic parameter estimates contribute to a better **understanding of genetic mechanisms**, and they are essential for the **prediction of breeding values**. Moreover, they are used to predict the expected genetic **response to selection** by the breeders equation (presented later in the course). Genetic parameters of interest include: additive genetic variance, non-additive genetic variance, environmental variance, heritability, genetic and phenotypic correlation, and repeatability. Genetic parameters are estimated using information on phenotypes and genetic relationships among individuals in the breeding population. In this section we will illustrate how different phenotypic sources and genetic relationships are used for estimating genetic parameters.

1.1 Genetic models and variance components

For estimating genetic parameters we need to specify a statistical model that describes the genetic and non-genetic factors that may affect the trait phenotypes. Often the non-genetic factors are described by systematic effects such as age, parity, litter size, days open, sex, herd, year, season, management, etc. (in animals) and sowing date, flowering time, location, block, etc. (in plants):

$$\text{phenotype} = \text{mean} + \text{systematic effect} + \text{genetic effect} + \text{residual}$$

Here we make a distinction between fixed effects, that determine the level (expected mean) of observations, and random effects that determine their variance. A model consists of at least one fixed effect (i.e. mean) and one random effect (the residual error variance). If observations also are influenced by a genetic contribution of the individuals, then a genetic variance component exists as well.

In general the total genetic effect for an individual is the sum of both additive and non-additive effects. However, only the additive genetic effects are passed on to the offspring and therefore contribute to the breeding value. Therefore we only consider the additive genetic model as the basis for estimation of genetic parameters. In that situation, **two components contribute to the total variance of observations σ_y^2 :** an additive genetic variance σ_a^2 and a residual variance σ_e^2 , which quantifies the variation of non-additive genetic effects as well as environmental deviations.

The statistical model is a formal representation of our quantitative genetic theory, but it is important to realize that all models are simple approximations of how genetic and non-genetic factors influence a trait. The goal of the statistical analysis is to find the best practical model that explains the most variation in the data. Statistical knowledge is required. The methods used for estimating genetic parameters is based on statistical concepts such as random variables, multivariate normal theory and linear (mixed) models. These concepts and their use will be explained in the following sections.

1.2 Genetic parameters

Heritability and genetic correlation are the key genetic parameters used in animal and plant breeding. They are defined in terms of the variance components (σ_a^2 and σ_e^2) defined in the previous section.

Heritability estimates the degree of variation in a phenotypic trait in a population that is due to genetic variation between individuals in that population. It measures how much of the variation of a trait can be attributed to variation of additive genetic factors, as opposed to variation of environmental factors and non-additive genetic factors. The **narrow-sense heritability** is the ratio of additive genetic variance (σ_a^2) to the overall phenotypic variance ($\sigma_y^2 = \sigma_a^2 + \sigma_e^2$):

$$h^2 = \sigma_a^2 / (\sigma_a^2 + \sigma_e^2) \quad (1)$$

A heritability of 0 implies that no additive genetic effects influence the observed variation in the trait, while a heritability of 1 implies that all of the variation in the trait is explained by the additive genetic effects. In general the amount of information provided by the phenotype about the breeding value is determined by the narrow-sense heritability. Note that heritability is population-specific and a heritability of 0 does not necessarily imply that there is no genetic determinism for the trait. The trait might be highly influenced by genetic factors (e.g., the number of arms in a human population). Yet, the observed variation for the trait might not be due to genetic factors, because all alleles contributing to the trait are fixed, and there are no segregating causal alleles for the trait, in the population (e.g., all people have two arms in this population). Therefore, observed variation may only be due to environmental factors (e.g., accidents resulting in severed arms), and the heritability in that population is then 0.

Genetic correlation is the proportion of variance that two traits share due to genetic causes. Genetic correlations are not the same as heritability, as it is about the overlap between the two sets of influences and not the absolute magnitude of their respective genetic effects; two traits could be both highly heritable but not genetically correlated, or they could have small heritabilities and be completely correlated (as long as the heritabilities are non-zero). Genetic correlation (ρ_a) is the genetic covariance between two traits divided by

the product of genetic standard deviation for each of the traits:

$$\rho_{a_{12}} = \frac{\sigma_{a_{12}}}{\sqrt{\sigma_{a_1}^2 \sigma_{a_2}^2}} \quad (2)$$

where $\sigma_{a_{12}}$ is the genetic covariance; $\sigma_{a_1}^2$ and $\sigma_{a_2}^2$ are the variances of the additive genetic values for the two traits in the population. A genetic correlation of 0 implies that the genetic effects on one trait are independent of the other, while a correlation of 1 implies that all of the genetic influences on the two traits are identical. Thus in order to estimate the heritability and genetic correlation we need to estimate the variance component defined above (σ_a^2 and σ_e^2), for each trait, in addition to the genetic covariance between traits.

1.3 Data required for estimating genetic parameters

Information on phenotypes and genetic relationships for individuals in a breeding population are, in combination with an appropriate statistical model, used for accurate estimation of genetic parameters and breeding values of individuals.

Phenotypes for traits of economic importance need to be recorded accurately and completely. All individuals within a production unit (herd, flock, ranch, plot) should be recorded, as sampling bias results in samples that are not representative of the population under study. Individuals should not be selectively recorded. Data includes the dates of events when traits are observed, factors that could influence an individual's performance, and an identification of contemporaries that are raised and observed in the same environment under the same management regime. Observations should be objectively measured, if at all possible.

Genetic relationships for the individuals in the breeding population are required. Genetic relationships can be inferred from a pedigree or, alternatively, computed from genomic data (genetic markers or DNA sequencing). Individuals and their parents need to be uniquely identified in the data.

Information about development (e.g., birth dates in animals, sowing date and flowering time in plants), breed composition, and genotypes for various markers could also be stored. If individuals are not uniquely identified, then genetic analysis of the population may not be possible at the individual level. In aquaculture species, for example, individual identification may not be feasible, but family identification (father and mother) may be known.

Prior information about the traits is useful. Read the literature. Most likely other researchers or breeders have already made analyses of the same species and traits. Even though their study populations are not the same as yours, their models could be useful starting points for further analyses. Their parameter estimates could result in useful predictions.

2 Methods for estimation of genetic parameters

In general, the estimation of heritability and genetic correlation relies on the resemblance between genetically related individuals. Close relatives share more alleles and, if the trait is under genetic influence, they will therefore share phenotypic similarities. Here we will present three methods for estimating heritability: parent-offspring regression, analysis of variance (ANOVA) for family data (e.g., half-sib/full-sib families) and restricted maximum likelihood (REML) analysis for a general pedigree. These methods are increasingly more complex, but they are also increasingly more flexible. While REML can analyze any type of relationships and structures, ANOVA can only analyze groups of individuals with similar relationships (e.g., half-sib or full-sib families), and regression analysis can only analyze pairs of individuals with similar relationships (e.g., pairs of parent and respective offspring, or pairs of monozygotic twins). These methods can also be used for estimation of genetic correlation, but this will not be covered in these notes.

2.1 Parent-offspring regression on pairs of observations

The simplest method for estimating genetic parameters is based on regression analysis. Heritability may be estimated by comparing phenotypes for traits recorded in parent and offspring. Parent-offspring regression

compares trait values in parents (y_p) to trait values in their offspring (y_o). Estimation of heritability is based on a linear regression model:

$$y_o = y_p b_{o|p} + e_o.$$

The slope of the regression line ($b_{o|p}$) is used to estimate the heritability of the trait when offspring values are regressed against the average phenotypic value of the parents (mid-parent regression) or the phenotypic values of one of the parents (single-parent regression). If only one parent's value is used then heritability is twice the slope. Therefore, the expected value of the regression line is $b_{o|p} = 0.5h^2$ (in single-parent regression) or h^2 (in mid-parent regression). Similar relationships can be derived for other types of parent-offspring regressions (mother-offspring or mid-parent-offspring regression). The heritability can therefore be estimated from the regression coefficient based on:

$$\begin{aligned} h^2 &= 2b_{o|m} && \text{(mother-offspring regression)} \\ h^2 &= 2b_{o|f} && \text{(father-offspring regression)} \\ h^2 &= b_{o|mf} && \text{(mid-parent-offspring regression)} \end{aligned}$$

Parent-offspring regression is not often used in practice. It requires data on two generations, and uses only this data. It is based on the genetic relationship between parent and offspring, but it is not possible to utilize genetic relationships among parents. However, the method is robust to selection of parents.

2.2 ANOVA on family data

Genetic parameters have been estimated for many years using analysis of variance (ANOVA). This method requires that individuals be assigned to groups with the same degree of genetic relationship. These are often half-sib groups, consisting of individuals with one parent in common (in animals, paternal half-sibs; in plants, maternal half-sibs), or full-sib groups, consisting of siblings with the same parents. In the case of half-sib groups, all offspring of one parent are treated as one group and offspring from different parents are allocated to different groups.

The ANOVA method has several limitations. First, we must assume that parents and families are unrelated. Second, data arising from experimental designs used for estimating genetic parameters are usually not balanced (i.e., number of offspring varies across families). Violations of these assumptions and unbalanced data will lead to biases or errors in the estimation of genetic parameters by the ANOVA method. Accommodations for imbalance (differences in number of offspring per family) are possible in ANOVA, but they are not straightforward. Moreover, imbalance is better accounted for by Restricted Maximum Likelihood approaches, which are also more flexible in genetic analyses.

2.3 Restricted Maximum Likelihood on general pedigree

Genetic parameters are nowadays estimated using restricted maximum likelihood (REML) or Bayesian methods. These methods allow for estimation of genetic parameters by information on the genetic similarity among individuals, from a general pedigree (with arbitrary relationships among them) or genomic data. This method allows for unbalanced data and accounts for genetic relationships within and between families. REML is based on linear mixed model methodology and uses a likelihood approach to estimate genetic parameters.

The linear mixed model contains the observation vector for the trait(s) of interest \mathbf{y} , its fixed effects that explain systematic differences in \mathbf{y} , and the random genetic effects \mathbf{a} and random residual effects \mathbf{e} . A matrix formulation of this general model equation is:

$$\mathbf{y} = \mu + \mathbf{a} + \mathbf{e} \tag{3}$$

where

- \mathbf{y} : is the vector of observed values of the trait,
- μ : is the population mean (representing the fixed effects),
- \mathbf{a} : is a vector of random genetic effects,
- \mathbf{e} : is a vector of random residual effects.

In the statistical model specified above, the random effects \mathbf{a} and \mathbf{e} and the phenotypes \mathbf{y} are random variables which follow a multivariate normal distribution:

$$\begin{aligned}\mathbf{a} &\sim N(\mathbf{0}, \mathbf{G}) \\ \mathbf{e} &\sim N(\mathbf{0}, \mathbf{R}) \\ \mathbf{y} &\sim N(\mu, \mathbf{V})\end{aligned}$$

where $\mathbf{G} = \mathbf{A}\sigma_a^2$, and $\mathbf{R} = \mathbf{I}\sigma_e^2$ are square matrices of genetic and residual covariance among the individuals, respectively, and $\mathbf{V} = \mathbf{A}\sigma_a^2 + \mathbf{I}\sigma_e^2$ is the overall phenotypic covariance matrix. \mathbf{A} is the additive genetic relationship matrix (which conveys similarity among additive genetic effects, depending on kinship), σ_a^2 is the additive genetic variance, \mathbf{I} is the identity matrix (which conveys independence among residuals) and σ_e^2 is the residual variance.

REML estimates the variance components σ_a^2 and σ_e^2 in the linear mixed model specified above. The general principle used in this is to find the set of parameters which maximizes the **likelihood** of the data, i.e., the probability of observations given the model and its parameter estimates: $p(\mathbf{y}|\hat{\mu}, \hat{\sigma}_a^2, \hat{\sigma}_e^2)$. From the REML estimate of the variance components, the heritability can easily be computed by

$$\hat{h}^2 = \hat{\sigma}_a^2 / (\hat{\sigma}_a^2 + \hat{\sigma}_e^2) \quad (4)$$

where ‘~’ refers to the estimate of the corresponding parameter.

2.3.1 Additive genetic relationships among individuals

Estimating heritability using REML (similarly to the parent-offspring regression and ANOVA method) requires that the phenotypic covariance between related individuals can be expressed by their genetic relationship and the genetic variance (σ_a^2). Related individuals share more alleles and thus resemble each other (i.e., they have correlated phenotypes, to an extent that depends on additive genetic relationships).

In general, the genetic covariance with a selection candidate (individual to be evaluated) depends on additive genetic relationships in matrix \mathbf{A} . Examples of additive genetic relationships A_{ij} between different individuals j and the selection candidate i (assuming no inbreeding) can be found in the table below.

Relative	A_{ij}
Self	1.0
Unrelated	0
Mother	0.5
Father	0.5
Grandparent	0.25
Half-sib	0.25
Full-sib	0.5
Cousin	0.0625
Progeny	0.5
Monozygotic twin	1
Dizygotic twin	0.5

The **A** matrix expresses the additive genetic relationship among individuals in a population, and is called the **additive relationship matrix**. The matrix **A** is symmetric and its diagonal elements A_{ii} are equal to $1 + F_i$, where F_i is the **coefficient of inbreeding** of individual i . F_i is defined as the probability that two alleles taken at random from individual i are identical by descent. As such, F_i is also the kinship coefficient of its parents.

Each off-diagonal elements A_{ij} is the genetic relationship between individuals i and j . In diploid individuals, A_{ij} twice the kinship coefficient between i and j , because each locus exists in two copies. Multiplying matrix **A** by the additive genetic variance σ_a^2 leads to the covariance among breeding values. Thus if a_i is the breeding value of individual i then

$$\begin{aligned} \text{Var}(a_i) &= A_{ii}\sigma_a^2 = (1 + F_i)\sigma_a^2 \\ \text{Cov}(a_i, a_j) &= A_{ij}\sigma_a^2 \end{aligned}$$

2.3.2 Algorithm to compute the numerator relationship matrix **A**

The matrix **A** can be computed using a recursive method. This method is especially suitable for an implementation by a software program. In what follows the recursive method to compute the elements of **A** is described. Initially, individuals in a pedigree are numbered from 1 to n and ordered such that parents precede their progeny. The following rules are then used to compute the elements of **A**.

- If both parents s and d of individual i are known, then
 - the diagonal element A_{ii} corresponds to: $A_{ii} = 1 + F_i = 1 + \frac{1}{2}A_{sd}$ and
 - the off-diagonal element A_{ji} is computed as: $A_{ji} = \frac{1}{2}(A_{js} + A_{jd})$
 - because **A** is symmetric $A_{ji} = A_{ij}$
- If only one parent s of individual i is known and assumed unrelated to its mate d
 - $A_{ii} = 1$
 - $A_{ij} = A_{ji} = \frac{1}{2}A_{js}$
- If both parents are unknown
 - $A_{ii} = 1$
 - $A_{ij} = A_{ji} = 0$

For example, we are given the following pedigree and we want to compute the matrix **A** using the recursive method described above.

Table 1: Example Pedigree To Compute Additive Genetic Relationship Matrix

	Calf	Sire	Dam
	3	1	2
	4	1	NA
	5	4	3
	6	5	2

The first step for the computations of **A** are the numbering and ordering of all the individuals. This is already done in the pedigree shown in Table 1. The elements of **A** are computed row-by-row starting with A_{11} .

$$\begin{aligned}
A_{11} &= 1 + F_1 = 1 + 0 = 1 \\
A_{12} &= 0 = A_{21} \\
A_{13} &= \frac{1}{2}(A_{11} + A_{12}) = 0.5 = A_{31} \\
A_{14} &= \frac{1}{2}A_{11} = 0.5 = A_{41} \\
A_{15} &= \frac{1}{2}(A_{14} + A_{13}) = 0.5 = A_{51} \\
A_{16} &= \frac{1}{2}(A_{15} + A_{12}) = 0.25
\end{aligned}$$

The same computations are also done for all the other elements of \mathbf{A} . The final result for the matrix looks as follows

$$\begin{bmatrix}
1 & 0 & 0.5 & 0.5 & 0.5 & 0.25 \\
0 & 1 & 0.5 & 0 & 0.25 & 0.625 \\
0.5 & 0.5 & 1 & 0.25 & 0.625 & 0.5625 \\
0.5 & 0 & 0.25 & 1 & 0.625 & 0.3125 \\
0.5 & 0.25 & 0.625 & 0.625 & 1.125 & 0.6875 \\
0.25 & 0.625 & 0.5625 & 0.3125 & 0.6875 & 1.125
\end{bmatrix}$$

As a result, we can see from the elements of the above shown matrix \mathbf{A} that individuals 1 and 2 are not related to each other. Furthermore from its diagonal elements, it follows that individuals 5 and 6 are inbred while individuals 1 to 4 are not inbred. Finally, we can see that different types of relationships were included in this data. In comparison, only two types of relationships could exist in regression and ANOVA analyses: unrelated (e.g., $A_{ij}=0$ between individuals from different families) or not (e.g., $A_{ij}=0.5$ between individuals from the same full-sib family).

2.3.3 Advantages of using REML for estimating genetic parameters

REML requires that \mathbf{y} has a multivariate normal distribution, although various authors have indicated that REML estimators may be an appropriate choice even if normality does not hold.

REML can account for selection (biased sampling of individuals from a population) when the complete mixed model is used with all genetic relationships and all data used for selection is included.

There is obviously an advantage of REML methods that are more flexible in handling animal and plant breeding data on several (overlapping) generations (and possibly several random effects). However, the use of such methods are “dangerous” in the sense we no longer need to think explicitly about the data structure. For example, to estimate additive genetic variance, we need to have a data set that contains a certain family structure which allows us to separate differences between families from differences within families. Or in other words, we need to differentiate genetic and residual effects, so the structure due to genetic relationships must be different from the structure due to residual effects (i.e., the G and R matrices must be different enough). In comparison ANOVA methods require more explicit knowledge about such structure, since the data has to be ordered according to family structures (e.g. by half-sib groups).

Early REML applications were generally limited to models largely equivalent to those in corresponding ANOVA analyses, considering one random effect only and estimating genetic variances from paternal half-sib covariances (so-called sire model in animal breeding). Today, heritability can be estimated based on genetic relationships, inferred from general pedigrees or estimated from genomic data. Linear mixed models are also used in genetic evaluation, allowing information on all known relationships between individuals to be incorporated simultaneously in the analysis.

Linear mixed models can include additional effects to describe the data more accurately: maternal, permanent environmental, cytoplasmic or dominance effects and QTL effects. These effects may be fitted as additional fixed or random effects.

3 When to estimate variance components?

In general, the estimation of variance components has to be based on sufficient data. Depending on the data structure and measurements, estimations can be based on hundreds (in selection experiments) or more than 10,000 observations (in field recorded data). In cases where the data set is small, the information from the literature may yield more accurate estimates of variance components (especially if the samples used in the literature are similar to our sample).

In general, we should estimate variance without external information if we study a new trait (for which no prior parameter estimates are available), or a different sample in which variances and covariances have changed over time due to various evolutionary forces (genetic drift, selection, migration, or mutation). It is well known that the genetic variance changes as a consequence of selection or genetic drift. Changes are expected, especially when generation intervals are short, selection intensity is high, or the trait under selection is determined by few causal genes with large effects. Moreover, the circumstances under which measurements are taken can change. If measurement conditions are better controlled, the environmental variance will decrease, and consequently the heritability will increase. Finally, the biological basis of a trait may change from one environment to another; for example, ‘feed intake under limited feeding’ is not really the same trait as ‘feed intake under ad-lib feeding’ (with potentially different causal loci and genetic effects). In conclusion, there are compelling reasons for regular estimation of genetic parameters on each new sample.

References

Michael Lynch and Bruce Walsh. *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Inc., Sunderland, USA, 1998. ISBN 0-87893-481-2.

Estimation of Breeding Values

Guillaume Ramstein & Peter Sørensen

2023-01-15

Contents

Learning objective:	1
1 Introduction	1
2 Fundamental concepts in breeding value estimation	2
2.1 Genetic model	2
2.2 Expected breeding value conditional on observed phenotypes	3
2.3 Accuracy of breeding value estimates	3
3 Breeding value estimation by phenotypic and genetic information	4
3.1 Own phenotype	4
3.2 Family data	4
3.3 General pedigree (multiple sources of information)	5

Learning objective:

This section introduces the basic concepts of breeding value estimation such as:

- genetic model used to estimate breeding values
- use of genetic parameters for estimating breeding values
- use of genetic relationships for estimating breeding values
- accuracy of estimated breeding values
- different methods, data sources and experimental designs for estimating breeding values

1 Introduction

Breeding value estimation is a fundamental component of breeding programmes, in which the breeding value of each individual is predicted to inform subsequent selection decisions. The breeding value for an individual is the additive genetic value which is passed to its offspring. Thus the breeding value is not a measure of how good an individual is in itself, but rather a measure of how good its offspring is expected to be (by the average effects of its genes in the population). Breeding values are used for:

- comparing individuals in the breeding population and selecting parents for the next generation
- predicting the consequences of selection decisions
- describing genetic differences over time (result of previous selection)

The true breeding values of individuals cannot be observed. It is only possible to measure their phenotypic values, which is influenced both by the genotype and the environment. Therefore, we must select individuals based on estimated breeding values.

2 Fundamental concepts in breeding value estimation

Breeding values are estimated using information on phenotypes and genetic relationships for individuals in a breeding population. As introduced previously the phenotype for a quantitative trait is the sum of both genetic and environmental factors. In general, the amount of information provided by the phenotype about the breeding value is determined by the narrow-sense heritability (h^2), which measures the proportion of additive genetic variance contained in the total phenotypic variance. Furthermore phenotypes collected from close relatives provide additional information about the breeding value of an individual. In this section we will show how phenotypic data and genetic relationships are used to estimate breeding values.

2.1 Genetic model

The true breeding value for an individual is the sum of all additive genetic effects that affect the quantitative trait:

$$a = \sum_{k=1}^K a_k$$

where a is the true breeding value of an individual, a_k is its additive genetic value at loci k , and K is the total number of causal loci. Because K is usually large, we assume (based on the central limit theory) that the true breeding value a is normally distributed:

$$a \sim N(0, \sigma_a^2)$$

Estimations of genetic parameters and breeding values are intricately related. Naturally, for breeding value estimation we will assume the same model used for estimation of genetic parameters (presented in the previous chapter):

$$\mathbf{y} = \mu + \mathbf{a} + \mathbf{e} \quad (1)$$

where

- \mathbf{y} is the vector of observed values of the trait,
- μ is the population mean (representing the fixed effects),
- \mathbf{a} is the vector of breeding values,
- \mathbf{e} is a vector of random residual effects.

As explained in the previous chapter, the random effects \mathbf{a} and \mathbf{e} and the phenotypes \mathbf{y} are random variables which follow multivariate normal distributions:

$$\begin{aligned}\mathbf{a} &\sim N(\mathbf{0}, \mathbf{A}\sigma_a^2) \\ \mathbf{e} &\sim N(\mathbf{0}, \mathbf{I}\sigma_e^2) \\ \mathbf{y} &\sim N(\mu, \mathbf{A}\sigma_a^2 + \mathbf{I}\sigma_e^2)\end{aligned}$$

where

- \mathbf{A} is the additive relationship matrix,
- σ_a^2 is the additive genetic variance,
- \mathbf{I} is the identity matrix, which conveys independence among residuals,
- σ_e^2 is the residual variance.

2.2 Expected breeding value conditional on observed phenotypes

The breeding value cannot be observed but must be estimated from phenotypic data and genetic relationships between individuals from the breeding population. Estimation of an unknown parameter using statistical modelling expresses the estimated quantity as a mathematical function of the observed data. The question is how this function should look like and what properties the estimated breeding values should have. For breeding purposes one objective for the estimated breeding values is that the response to selection is maximized. Henderson (1963) found that the improvement of an offspring generation compared to the parent generation can be maximized when parents are selected based on the conditional expected values ($E(\mathbf{a}|\mathbf{y})$) of the true breeding values \mathbf{a} given the observed phenotypic values \mathbf{y} . Under the assumption of multivariate normality (see previous section), the expected value of the breeding value conditional on the observed phenotype y is written as:

$$E(\mathbf{a}|\mathbf{y}) = E(\mathbf{a}) + b_{\mathbf{a}|\mathbf{y}}(\mathbf{y} - E(\mathbf{y}))$$

The breeding value is defined as deviation from the general mean, which means that the expected value $E(\mathbf{a})$ of the true breeding value is 0. Therefore the expected value of the breeding value is:

$$E(\mathbf{a}|\mathbf{y}) = b_{\mathbf{a}|\mathbf{y}}(\mathbf{y} - \mu) \quad (2)$$

The expression for the estimate of the breeding value consists of two parts:

- $\mathbf{y} - \mu$: the deviation of observed values from the population mean,
- $b_{\mathbf{a}|\mathbf{y}}$: the regression coefficient which relates that phenotypic deviation to the breeding value.

To be able to estimate the breeding value we need to estimate μ , $Var(\mathbf{y})$, and $Cov(\mathbf{a}, \mathbf{y})$ in the expression above. It is possible to derive simple formulas for these terms based on:

- adjusted phenotypic observations for the quantitative trait of related individuals: μ
- heritability of the quantitative trait: \hat{h}^2
- genetic relationship among individuals: \mathbf{A}

We will distinguish true and estimated breeding value using the following notation:

- \mathbf{a} : additive genetic value = true breeding value
 $\hat{\mathbf{a}} = E(\mathbf{a}|\mathbf{y})$: estimated additive genetic value = estimated breeding value

2.3 Accuracy of breeding value estimates

Estimated breeding values $\hat{\mathbf{a}}$ are estimates of the true breeding values \mathbf{a} , which cannot be observed directly. It is important to determine how well we have estimated the breeding value in relation to the true breeding value. This can be done using accuracy or reliability.

For each individual, the **accuracy** is the correlation between estimated and true breeding value, $r_{a,\hat{a}}$. A high correlation means that the estimated breeding value is accurate, so that we are confident in making a selection based on \hat{a} .

The **reliability** is just the squared accuracy: $r_{a,\hat{a}}^2$. It can be shown that the reliability is the ratio of the variance of estimate breeding value, over the variance of true breeding value:

$$r_{a,\hat{a}}^2 = \frac{Var(\hat{a})}{\sigma_a^2}$$

Therefore, the reliability can be interpreted as the part of the genetic variation that is explained by the estimated breeding values. The reliability of the breeding value estimates is important because it determines

how well we can predict an individual's genetic value. It can be used to control the risk of a breeding plan: for example, a low $r_{a,\hat{a}}^2$ leads to greater "risk" for both lower and higher true breeding value and we might consider adding phenotypic records, in order to make better-informed selection decisions. Lastly reliability is one of the crucial factors that determine genetic progress, according to the breeder's equation (which will be introduced later in the course).

3 Breeding value estimation by phenotypic and genetic information

We will illustrate the basic principles of breeding value estimation using some simple examples where the trait has been measured on the individuals themselves or close relatives.

3.1 Own phenotype

The breeding value (a) is estimated from an individual's own phenotype (y) by:

$$\hat{a} = h^2(y - \mu)$$

Thus the estimated breeding values based on own phenotypic records can be computed by the estimated trait heritability (h^2) and the observed phenotype deviation ($y - \mu$).

The accuracy for the breeding value based on own phenotype (y) is calculated as:

$$r_{a,\hat{a}} = \sqrt{h^2}$$

The use of records on the candidate itself is called **performance testing**. Importantly, if the heritability is the same for all individuals, then selection by performance testing is equivalent to selection by observed phenotypes (individuals' ranking will be exactly the same)! However, weighting by the heritability allows us to accommodate differences in h^2 across individuals. In general, using equation (2) allows us to account for differences in the amount of genetic information between individuals. Such differences will be more important as we consider more general data types (family data, or general pedigrees).

Estimation of breeding value based on own phenotype is only possible when the trait in question can be measured (directly or indirectly) on the breeding individual, i.e., the candidate to be evaluated for selection. Sometimes this is not possible, e.g., traits that are sex-limited (milk production, female fertility, etc.) cannot be measured in male individuals. Traits like carcass composition and meat quality cannot be measured on live animals, unless an indirect method can be used (e.g. ultra-sonic measurement of carcass composition). In this situation it might be possible to use phenotypic information on relatives.

3.2 Family data

In practice we often use phenotypic records from close relatives, such as progenies, half-sibs, full-sibs, parents and grandparents. Phenotypes collected on close relatives (as compared to distant relatives) provide more information about the breeding value of an individual (as close relatives share more DNA in common).

General formula for estimating breeding values: In family data, individuals are evaluated through phenotypic records at their relatives: for example, the breeding value of each bull is estimated by observed phenotypes at its daughters. By applying formula (2) to family means \bar{y} (the average of observed phenotypes in each family), we obtain:

$$\hat{a} = b_{a|\bar{y}}(\bar{y} - \mu)$$

where

$$b_{a|\bar{y}} = \frac{A'n h^2}{1 + (n - 1)r}$$

where A' is the genetic relationship between the breeding individual and relatives, n is the number of relatives per family, h^2 is the trait heritability, and r is correlation between relatives within each family ($r = A'' h^2$, where A'' is the genetic relationship among relatives). Here we assume that individuals do not share any common environment, even though in practice they will certainly have similar upbringing (e.g., same litter in animals, or same plot in plants).

Thus the importance given to a specific source of information depends on the additive genetic relationships (A' and A''), the heritability of the trait (h^2), and the amount of phenotypic information in each family (n), i.e. the number of relatives (progenies, sibs, etc.).

In the case of **progeny testing**, the breeding value of a selection candidate is estimated by the average of phenotypic records in its progeny. Then, $A' = \frac{1}{2}$ (relationship between the selection candidate and each of its progeny) and $A'' = \frac{1}{4}$ if the progeny are half-sibs, or $A'' = \frac{1}{2}$ if they are full-sibs.

General formula for reliability of estimated breeding value: The accuracy for the breeding value based on a family mean (\bar{y}) is calculated as:

$$r_{a,\hat{a}}^2 = \frac{(A')^2 nh^2}{1 + (n - 1)r}$$

Thus reliability depends on the same factors as the estimated breeding value except for the phenotypic value. Although the reliability depends on the number of records it does not depend on the numerical value of phenotypes. From this formula is it clear that higher reliability (and accuracy) can be achieved when:

- genetic relationship to individuals with information (A') is high
- there are many records (n is high)
- heritability (h^2) is high
- correlation between records ($r = A'' h^2$) is low (little redundancy in observations)

Phenotypic records on progenies are generally the most accurate source of information for genetic evaluation (high genetic relationship A' and high n). Progeny testing is especially useful when the heritability is low. For example, it can be much more accurate than evaluation on own phenotype when the heritability is low (say, 0.1) and the progeny is large (~100-150). The main disadvantage of progeny testing is that it takes resources (time and money) before results on progenies are available. In animal breeding, progeny testing is often used for male animals as they usually get many more progeny (larger n) than females, especially when artificial insemination is practiced.

3.3 General pedigree (multiple sources of information)

So far, we have considered methods that are instructive, because their formulas point to the factors contributing to accurate estimation of breeding values (heritability, family size, etc.). However, they are not flexible, because they require fixed relationships (e.g., pairs of parent and offspring, or family of half-sibs). In practice, it is common to combine information from individuals from a **general pedigree**, in which there are varying degrees of relationships among individuals. All information available is then utilized to estimate the breeding value of each candidate.

Phenotypic records on the candidate's sibs, half-sibs and full-sibs, are often used in addition to the candidate's own record, or as surrogate measurements, for example on traits that cannot be measured on the candidate itself. Common-environment effects (e.g. full-sibs raised in the same herd) may bias the estimation of breeding values, unless we are able to adjust for them (e.g., by additional fixed effects in linear models).

Parental information at different generations (parents, grandparents, etc.) is generally available even before the candidate is born, and can provide information very early. However, the additive genetic relationship, and thus the proportion of common genes between the candidate and the pedigree, is halved for every generation backwards (at least 0.5 for a parent, 0.25 for a grandparent, etc.). Moreover, there is redundancy in the information provided by different generations of parents. For example, if there are accurate estimate of the parents' breeding values, then there is little to gain in using information on grandparents (actually, if the parents true breeding values are known, there would be no additional gain of information from grandparents).

3.3.1 BLUP, a general approach for estimation of breeding values

Compared to the idealized cases described in the previous section, a practical breeding scenario poses two problems: accounting for heterogeneous sources of genetic information (different types of relatives); and adjusting for fixed effects in the breeding population(s) (fixed environmental or genetic effects). The solution to these problems was presented by Charles R. Henderson, as the **BLUP** method. The properties of the BLUP methodology developed by Henderson are directly incorporated into its name:

- **B** stands for **best** which means that the correlation between the true (a) and the predicted breeding value (\hat{a}) is maximal or the prediction error variance ($Var(a - \hat{a})$) is minimal.
- **L** stands for **linear** which means the predicted breeding values are linear functions of the observations (y).
- **U** stands for **unbiased** which means that the expected values of the predicted breeding values are equal to the true breeding values.
- **P** stands for **prediction**.

The estimation of breeding values by BLUP corrects for the redundancy between different sources (e.g., the redundant information provided by parents and grandparents). Moreover, they adjusted for fixed effects in the populations. So far we have referred to that fixed effect as the population mean and we have assumed this adjustment μ to be known. Indeed, in model (1) we defined the true breeding values \mathbf{a} and the non-identifiable environmental effects \mathbf{e} as deviations from a common mean, the average effect of all fixed genetic and environmental factors captured by the population mean μ . But this is only true in a single idealized population where all selection candidates are kept in the same environment in which they deliver their performances at the same time. In practice, the phenotypic records often need to be adjusted for systematic (fixed) effects, such as age, parity, litter size, days open, sex, herd, year, season, management, etc. Several of those effects fluctuate very little over time, so accurate estimates of their effect may be obtained from previous (“historical”) data sets. Effects of factors like herd, year, season, and management fluctuate more, and are therefore best estimated directly from the data to be used in the genetic evaluations.

BLUP estimates are widely used in genetic evaluations, for both traditional estimation of breeding values and estimation of genomic breeding values (where \mathbf{A} is determined by genomic data; see next chapter). The theoretical foundations of BLUP estimates, and the development of efficient algorithms for computing them, together with the availability of large computational resources at a very low price, have made BLUP the de facto standard for breeding value estimation.

3.3.2 Estimating random genetic effects in a linear mixed model by BLUP

The goal of the BLUP analysis is the estimate the random genetic effects, \mathbf{a} , in the linear mixed model specified above (model (1)). This can be done using the BLUP equations shown below. BLUP estimates can be derived by applying formula (2) to general pedigree data. The BLUP estimate of $\hat{\mathbf{a}}$ is then:

$$\hat{\mathbf{a}} = \mathbf{A} \mathbf{V}^{-1} (\mathbf{y} - \hat{\mu}) \quad (3)$$

The BLUP equation for the estimation of breeding values consists of three parts;

- The term $\mathbf{y} - \hat{\mu}$ shows that the observed phenotypic values are corrected for the population mean represented by $\hat{\mu}$.
- The covariance between the true breeding values (\mathbf{a}) and phenotypes (\mathbf{y}) is $Cov(\mathbf{a}, \mathbf{y}) = \mathbf{A}$.
- The inverse of the phenotypic variance is $[Var(\mathbf{y})]^{-1} = \mathbf{V}^{-1}$, following (1).

\mathbf{V}^{-1} is the matrix inverse of the phenotypic variance matrix \mathbf{V} . It accounts for correlation among individuals, and redundancy in genetic information in the sample.

3.3.3 BLUP breeding values are optimal for ranking and selection

Selection on BLUP estimates of breeding values maximizes the probability for correct ranking of breeding individuals. Therefore, selection by BLUP maximizes genetic gain from one generation to another. There are

many factors that contribute to this advantage:

- The linear mixed model makes full use of information from all relatives
- The breeding values are adjusted for systematic environmental effects in an optimal way. This means that individuals can also be compared across herds, age classes, plots etc.
- Non-random mating is accounted for, by the information in \mathbf{A}
- Bias due to culling within generation (e.g., between the 1st and 2nd lactations in dairy cattle) and selection (over generations) is accounted for, assuming that also non-selected individuals' data are included in the analysis.

It should, however, be noted that the genetic evaluation is based on phenotypic observations, and that regardless of how great the BLUP procedure may be, it cannot compensate for bad data. So a good recording system is necessary for a reliable genetic evaluation and subsequent genetic gain. Also, BLUP does not account for the uncertainty about the estimation of genetic parameters, and assumes that the genetic parameter estimates used are the true ones. In practice that means that BLUP estimates will only be accurate if the estimated genetic parameters in model (1) are close enough to their true value.

It should be noted that there is a potential risk for increased inbreeding when selection is based on information from all relatives. The probability that several family members are selected jointly increases, which may result in increased inbreeding. To avoid this, and to optimize long-term selection response, selection on BLUP estimates might be combined with some restriction on average relationship among selected individuals.

References

Estimation of Genomic Breeding Values

Guillaume Ramstein & Peter Sørensen

2023-01-15

Contents

Learning objective:	1
1 Introduction	1
2 Genomic information	2
2.1 Genomic markers	2
2.2 Quantitative Trait Loci and Linkage Disequilibrium (LD)	2
3 Genomic BLUP (GBLUP)	3
3.1 General approach for estimating genomic breeding values	3
3.2 Linear mixed model in GBLUP	3
4 Genomic prediction by GBLUP	4
4.1 Genomic Relationship Matrix G	4
4.2 Computation of genomic estimated breeding values	4
4.3 Accuracy of genomic breeding values	4
5 Impact of genomic selection on breeding programmes	5

Learning objective:

This section introduces the basic concepts of estimating genomic breeding values such as:

- use of genomic relationships for estimating breeding values
- accuracy of genomic estimated breeding values
- methodological and practical comparison between genomic selection and traditional selection

1 Introduction

A new technology called **genomic selection** has revolutionized animal and plant breeding in the last twenty years. Genomic selection refers to selection decisions based on genomic estimated breeding values (GEBVs).

We have previously learned how phenotypic records and genetic relationships computed from pedigree information can be used to compute estimated breeding values (EBVs) by *pedigree-based* BLUP. Genome-wide DNA markers can replace or supplement pedigree information in **genomic prediction**. The concepts and methods of genomic prediction were introduced by Meuwissen et al. (2001). They showed that information from genotypes of many marker loci evenly spread over the complete genome can successfully be used to estimate genomic breeding values. Because the genomic marker data is spread over the complete genome it is often referred to as **genomic information**. The term of genomic selection was coined for the use of this information for selection purposes.

Genomic selection were not implemented until Schaeffer et al. (2006) showed that the introduction of genomic selection could save up to about 90% of total costs in a cattle breeding programme, provided the accuracies computed by Meuwissen et al. (2001) were achieved. After the publication of Schaeffer et al. (2006), many animal and plant breeding institutions introduced procedures for genomic selection.

2 Genomic information

In recent years much attention has been paid to genomic information, because of dramatic developments in genotyping technologies. Today, dense genetic maps, which provide information about genetic markers and their localization in the genome, are available for many important animal and plant species (Table 1). Genetic maps are still missing in several species, but this lack of genomic information can be circumvented by using RAD-sequencing (Restriction site associated DNA sequencing) or Genotyping-by-Sequencing (GBS), which enable dividing the entire genome into smaller segments for subsequent sequencing. Ultimately, genomes may be fully sequenced, without resorting to marker information. This is possible, but still expensive. So, relatively few individuals are routinely sequenced in breeding programmes (in animals, mostly bulls, horses, and dogs; in plants, founder accessions in major crops like maize, rice, barley, and wheat).

Table 1. Number of markers used for genomic selection in different species

Species	No. SNPs (in thousands)	Genome size ($\times 10^9$)
Cattle	778	2.67
Pig	64	2.81
Chicken	581	1.05
Horse	70	2.47
Sheep	54	2.62
Dog	170	2.41
Maize	*	2.3
Wheat	*	17.1
Rice	*	0.38
Barley	*	5.3
Tomato	*	0.83

SNP: single nucleotide polymorphisms *: Number of SNPs vary greatly depending on the assay: from about 6,000 to 900,000 in SNP arrays, to several millions in whole-genome sequencing data.

2.1 Genomic markers

The different polymorphisms in the genome that are used in genomic selection are called **markers**. The so-called **single nucleotide polymorphisms** (SNPs) have been shown to be the most useful types of markers. SNPs are single-base changes in the DNA at specific positions in the genome. According to empirical analyses, most SNPs have only two different states (e.g., A or T at a specific locus), although they could in theory have up to four different alleles (one for each possible DNA base).

It is important that SNPs be more or less evenly spread over the entire genome. Some SNPs may be located in protein-coding regions, but they should also be located in non-coding regions: gene regulatory regions or intergenic regions.

2.2 Quantitative Trait Loci and Linkage Disequilibrium (LD)

The loci that are relevant for a quantitative trait are called **Quantitative Trait Loci** (QTL). A SNP marker is informative about a QTL if it is in **linkage disequilibrium** (LD) with it, i.e., there is a non-random association between the QTL alleles and the SNP alleles. If a positive QTL allele evolved in the neighborhood of several SNP loci, the positive QTL allele is often co-inherited with the same SNP allele. Therefore, the SNP is said to *tag* the QTL of interest. Over generations, recombination between the QTL and the neighboring SNP locus may weaken the statistical association between the SNP and the QTL. This effect of recombination

is smaller when the QTL and the SNP locus are physically closer together on the same chromosome. For most animal and plant species (genome sizes from hundreds of millions to few billion bases), about 50,000 SNP markers are required to get a sufficient coverage of the complete genome, and ensure that most QTLs for any given trait is sufficiently tagged by at least one SNP.

3 Genomic BLUP (GBLUP)

3.1 General approach for estimating genomic breeding values

Genomic estimated breeding values are estimated from marker information and phenotypic records. First, the genotypes of individuals are called at marker loci in the genome. Second, the **marker effects** (additive allele effects) at all marker loci are estimated simultaneously. Finally, the **genomic estimated breeding value (GEBV)** of each individual is calculated as the sum of its alleles weighted by their respective marker effect.

Marker effects are estimated in a group of individuals called the **reference population**. For any other individual, only a blood or tissue sample is needed to compute its GEBV. For breeding purposes, it is desirable that the GEBV be estimated accurately, and early in the individual's life (before mating at the latest). Marker effects can be estimated if there are phenotypes and marker genotypes from many individuals (from several hundreds to hundreds of thousands or even millions). With sufficiently dense marker maps, marker effects capture the genomic variability in the population in which they were estimated, because markers are in tight LD with the causal loci that they bracket.

3.2 Linear mixed model in GBLUP

There exist different approaches for computing GEBVs, but the most commonly used is by far **genomic BLUP**. GBLUP adapts the BLUP method (introduced in the previous chapter) to genomic information: it uses **genomic relationships** instead of pedigree-based relationships to describe genetic similarities between individuals. Under the GBLUP approach, genetic parameters and GEBVs are estimated in the reference population, in a linear mixed model. GEBVs of selection candidates are then estimated based on their genomic relationship to the individuals in the reference population.

GBLUP allows for estimation of GEBVs for individuals without phenotypic records and close genetic relationships. This is the main advantages of genomic selection. As soon as DNA is available for an individual, its marker genotypes can be determined and a genomic breeding value can be estimated. Furthermore, GEBVs are generally more accurate than the traditional estimated breeding values (EBVs) based on pedigree information.

The linear mixed model used in GBLUP is similar to the linear mixed model we used in the last two chapters for estimating genetic parameters and breeding values based on pedigree information. The main difference between GBLUP and pedigree BLUP is that genetic similarities are determined by a **genomic relationship matrix \mathbf{G}** instead of a pedigree-based genetic relationship matrix **\mathbf{A}** :

$$\mathbf{y} = \mu + \mathbf{a} + \mathbf{e} \quad (1)$$

where μ , \mathbf{a} and \mathbf{e} are as introduced previously:

- \mathbf{y} is the vector of observed values of the trait,
- μ is the population mean (representing the fixed effects),
- \mathbf{a} is the vector of breeding values,
- \mathbf{e} is a vector of random residual effects.

Again, the random effects \mathbf{a} and \mathbf{e} , and the phenotypes \mathbf{y} are random variables which follow multivariate normal distributions. However, the covariance among breeding values in \mathbf{a} is now determined by **\mathbf{G}** :

$$\begin{aligned}\mathbf{a} &\sim N(\mathbf{0}, \mathbf{G}\sigma_a^2) \\ \mathbf{e} &\sim N(\mathbf{0}, \mathbf{I}\sigma_e^2) \\ \mathbf{y} &\sim N(\mu, \mathbf{G}\sigma_a^2 + \mathbf{I}\sigma_e^2)\end{aligned}$$

where

- \mathbf{G} is the genomic relationship matrix,
- σ_a^2 is the additive genomic variance,
- \mathbf{I} is the identity matrix, which conveys independence among residuals,
- σ_e^2 is the residual variance.

4 Genomic prediction by GBLUP

4.1 Genomic Relationship Matrix \mathbf{G}

The genomic relationship matrix \mathbf{G} is constructed using genomic markers, as follows:

$$\mathbf{G} = \frac{\mathbf{W}\mathbf{W}^T}{\sum_{k=1}^m 2p_k(1-p_k)} \quad (2)$$

where \mathbf{W} is the centered marker matrix, and m is the total number of markers. Each column in \mathbf{W} contains allelic information at a given marker, across individuals. At the k^{th} column, the marker data \mathbf{w}_k is calculated as follows: $\mathbf{w}_k = \mathbf{x}_k - 2p_k$, where p_k is the minor allele frequency at the k^{th} genomic marker and \mathbf{x}_k is the k^{th} column vector of an allele count matrix, which contains the genotypes coded as 0, 1 or 2 counting the number of minor alleles at all markers. The mean of allele counts at marker k is $2p_k$, so $\mathbf{x}_k - 2p_k$ is the vector of centered allele counts across individuals at marker k .

The centering of the allele counts and the scaling factor $\sum_{k=1}^m 2p_k(1-p_k)$ ensure that the genomic relationship matrix \mathbf{G} has similar properties as the pedigree-based relationship matrix \mathbf{A} . The main difference between the two types of relationship matrices is that \mathbf{A} is based on *expected* identity by descent (expected sharing of the same alleles, transmitted from common ancestors) whereas \mathbf{G} is based on *realized* identity by state (observed sharing of the same alleles, regardless of their origin).

4.2 Computation of genomic estimated breeding values

Similarly to EBVs, the GEBVs $\hat{\mathbf{a}}$ in the linear mixed model specified above can be computed by the following BLUP equation:

$$\hat{\mathbf{a}} = \mathbf{G}\mathbf{V}^{-1}(\mathbf{y} - \hat{\mu}) \quad (3)$$

From equation (3) we can see that the GBLUP approach is very similar to the pedigree-based breeding value estimation procedure. In GLUP, the covariances between breeding values is based on the genomic relationship matrix \mathbf{G} which is computed from genomic markers whereas, in pedigree-based BLUP, it is based on the relationship matrix \mathbf{A} computed from pedigree information.

4.3 Accuracy of genomic breeding values

GBLUP estimation is generally more accurate than pedigree-based BLUP estimation. One reason for this is that the genomic relationship matrix uses phenotypic information more efficiently. Indeed, it captures actual similarities among individuals (based on observed degree of allele sharing), rather than the expected similarities based on pedigree data which may be incomplete or erroneous.

Like the accuracy of EBVs, the accuracy of GEBVs depends on the trait heritability, the number of phenotypic records, and the relationships among individuals. The accuracy of genomic breeding values increases when the size of the reference population increases, provided that:

- the reference population represents as much of the relevant genetic variation in the breeding population as possible
- selection candidates are closely related to the reference population
- QTL for the trait are adequately tagged by the genomic markers

5 Impact of genomic selection on breeding programmes

The relative benefit of genomic selection depends on the efficiency of traditional breeding value estimation. If all selection candidates already have accurate EBVs at the time of selection, then genomic information may not add any economic value. Hence, genome-wide marker information is most useful when phenotypic recording is restricted – for instance when phenotypes are expressed late in the animal's life (e.g., meat quality, longevity), are expressed in only one sex (e.g., milk yield, egg production in animals; grain yield in dioecious plants), or are expensive to measure (e.g., feed efficiency, bacteriological samples, progesterone profiles or other physiological measurements in animals; metabolic and physiologic measurements in plants). Furthermore, genome-wide marker information is useful for traits with low heritability, provided a sufficient amount of phenotypes can be recorded. It should therefore be clear that the extra benefits of genomic information vary across traits and across species, although it can in principle be used for all species and traits.

Dairy cattle breeding is characterised by the main traits only being measurable in females while very intense selection is only possible in males. Thus genome-wide markers are very useful in dairy cattle breeding.

In pig breeding, most traits are measured on all selection candidates before sexual maturity. Therefore genomic information gives less extra value for pig breeding compared to dairy cattle. However, exceptions for pigs include litter size (only measurable in females and after sexual maturity), feed efficiency (only measured on few animals because it is expensive), longevity, and carcass traits (expressed late).

In crops, genomic selection is especially useful when the generation interval (the time between breeding cycles) is drastically reduced, e.g., from 6 years in traditional breeding to 1 year in genomic selection. Because evaluation of plant varieties across enough environments is time-consuming, genomic selection is potentially useful in most crop species. But, it is especially promising in perennial species, which require even more time for testing. Such species include trees for wood or fruits (e.g., poplar, apple), and grasses for forage, bioenergy or grain (e.g., perennial ryegrass, miscanthus, perennial wheat).

Genomic breeding values can be used to enhance the screening of potential breeding individuals (pre-selection), which is especially attractive in situations when the costs of genotyping are relatively low compared to the costs of recording phenotypes.

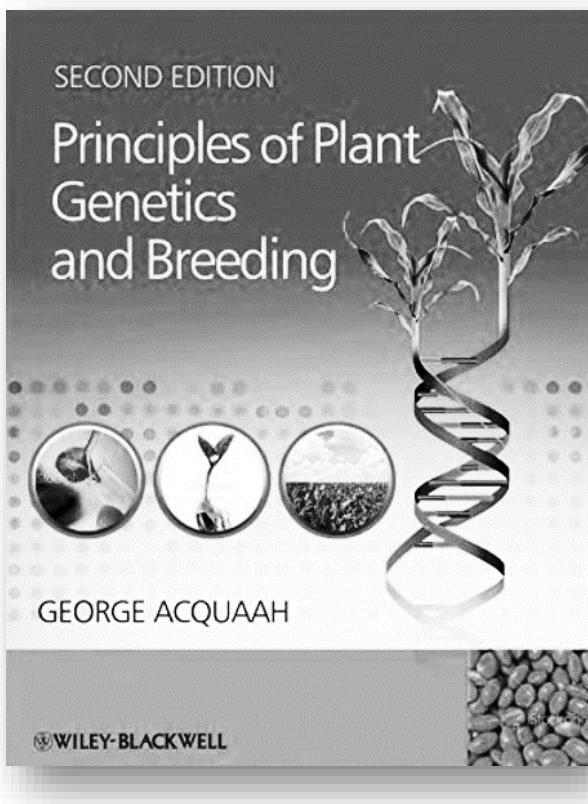
Genomic breeding values are also useful in intensifying selection for young animals, thereby facilitating a reduced overall generation interval. For instance, with the availability of accurate GEBVs for young bulls, it is more attractive to use the best young bulls widely rather than having to wait for the results of progeny testing. Here the substantial reduction in generation interval offsets the slightly lower accuracy of GEBVs compared to EBVs by progeny testing (most accurate but very time-consuming).

Another benefit of using GEBVs rather than traditional pedigree-based EBVs is that they result in less inbreeding if the same selection intensities are maintained. This happens because traditional pedigree-based breeding values put more emphasis on parent information than genomic breeding values, especially for traits with low heritability.

2. PLANTS



2.1 Study book



Title:	<i>Principles of Plant Genetics and Breeding</i>
Edition:	2
Authors:	George Acquaah
Year:	2012
Publisher:	John Wiley & Sons Inc

2.1.1-2.1.5

Acquaah (2012)

Page 97-120, 121-129, 173-183, 185-197, and 199-212



5

Introduction to reproduction and autogamy

Purpose and expected outcomes

Rudolph Camerarius is credited with establishing sexual differentiation, noting that male and female sex organs exist in the Plant Kingdom. Some species produce flowers while others do not. In flowering species, reproduction involves the union of gametes following pollination. Plant breeders need to understand the mode of reproduction in order to manipulate plants effectively to develop new and improved ones for crop production. After studying this chapter, the student should be able to:

- 1 Discuss the importance of the mode of reproduction to plant breeding.
- 2 Distinguish between self-pollination and cross-pollination.
- 3 Discuss the natural barriers that favor or hinder each of the modes of reproduction.
- 4 Discuss the implications of mode of reproduction in schemes and strategies employed in plant breeding.
- 5 Discuss the use of male sterility and self-incompatibility in breeding.

5.1 Importance of mode of reproduction to plant breeding

Plant breeders need to understand the reproductive systems of plants for the following key reasons:

- (i) The genetic structure of plants depends on their mode of reproduction. Methods of breeding are generally selected such that the natural genetic

structure of the species is retained in the cultivar. Otherwise, special efforts will be needed to maintain the newly developed cultivar in cultivation.

- (ii) In flowering species, artificial hybridization is needed to conduct genetic studies to understand the inheritance of traits of interest and for transfer of genes of interest from one parent to another. To accomplish this, the breeder needs

to understand thoroughly the floral biology and other factors associated with flowering in the species.

- (iii) Artificial hybridization requires an effective control of pollination so that only the desired pollen is allowed to be involved in the cross. To this end, the breeder needs to understand the reproductive behavior of the species. Pollination control is critical to the hybrid seed industry.
- (iv) The mode of reproduction also determines the procedures for multiplication and maintenance of cultivars developed by plant breeders.

5.2 Overview of reproductive options in plants

Four broad and contrasting pairs of reproductive mechanisms or options occur in plants.

- (i) **Hermaphrodity versus unisexuality.** Hermaphrodites have both male and female sexual organs, and hence may be capable of self-fertilization. On the other hand, unisexuals, having one kind of sexual organ, are compelled to cross-fertilize. Each mode of reproduction has genetic consequences. Hermaphrodity promotes a reduction in genetic variability, whereas unisexuality, through cross-fertilization, promotes genetic variability.
- (ii) **Self-pollination versus cross-pollination.** Hermaphrodites that are self-fertile may be self-pollinated or cross-pollinated. In terms of pollen donation, a species may be **autogamous** (pollen comes from the same flower – selfing) or **allogamous** (pollen comes from a different flower). There are finer differences in these types. For example, there may be differences between the time of pollen shed and stigma receptivity.
- (iii) **Self-fertilization versus cross-fertilization.** Just because a flower is successfully pollinated does not necessarily mean fertilization would occur. The mechanism of self-incompatibility causes some species to reject pollen from their own flowers, thereby promoting outcrossing.
- (iv) **Sexuality versus asexuality.** Sexually reproducing species are capable of providing seed through sexual means. Asexuality manifests in one of two ways – vegetative reproduction (in which no seed is produced) or agamospermy (in which seed is produced).

5.3 Types of reproduction

Plants are generally classified into two groups, based on the mode of reproduction, as either **sexually reproducing** or **asexually reproducing**. Sexually reproducing plants produce seed as the primary propagule. Seed is produced after sexual union (fertilization) involving the fusion of sex cells or **gametes**. Gametes are products of meiosis and, consequently, seeds are genetically variable. Asexual or vegetative reproduction mode entails the use of any vegetative part of the plant for propagation. Some plants produce modified parts, such as creeping stems (stolons or rhizomes), bulbs or corms, which are used for their propagation. Asexual reproduction is also applied to the condition whereby seed is produced without fusion of gametes (called apomixis). It should be pointed out that some plants could be reproduced by either the sexual or asexual mode. However, for either ease of propagation or for product quality, one mode of reproduction, often the vegetative mode, is preferred. Such is the case in flowering species such as potato (propagated by tubers or stem cuttings) and sugarcane (propagated by stem cuttings).

5.4 Sexual reproduction

Sexual reproduction increases genetic diversity through the involvement of meiosis. Flowering plants dominate the terrestrial species. Flowering plants may reproduce sexually or asexually.

5.4.1 Sexual lifecycle of a plant (alternation of generation)

The normal sexual lifecycle of a flowering plant may be simply described as consisting of events from first establishment to death (from seed to seed in seed-bearing species). A flowering plant goes through two basic growth phases – **vegetative** and **reproductive**, the former preceding the latter. In the vegetative phase, the plant produces vegetative growth only (stem, branches, leaves, etc., as applicable). In the reproductive phase, flowers are produced. In some species, exposure to a certain environmental factor (e.g., temperature, photoperiod) is required to switch from the vegetative to the reproductive phase. The duration between phases varies among species and

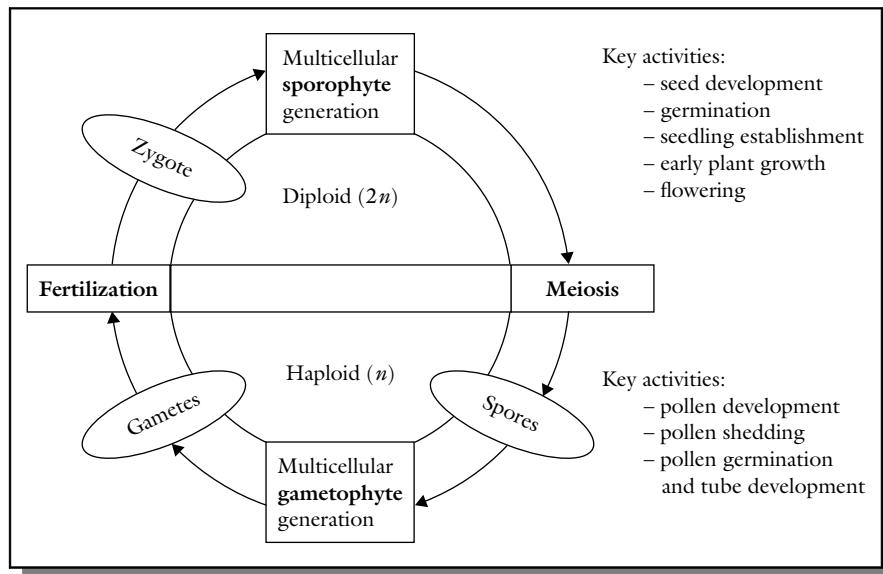


Figure 5.1 Schematic representation of the alternation of generations in flowering plants. The sporophyte generation is diploid and often the more conspicuous phase of the plant lifecycle. The gametophyte is haploid.

can be manipulated by modifying the growing environment.

For sexual reproduction to occur, two processes must occur in sexually reproducing species. The first process, meiosis, reduces the chromosome number of the diploid ($2n$) cell to the haploid (n) number. The second process, **fertilization**, unites the nuclei of two gametes, each with the haploid number of chromosomes, to form a diploid. In most plants, these processes divide the lifecycle of the plant into two distinct phases or generations, between which the plant alternates (called **alternation of generation**) (Figure 5.1). The first phase or generation, called the **gametophyte generation**, begins with a haploid spore produced by meiosis. Cells derived from the gametophyte by mitosis are haploid. The multicellular gametophyte produces gametes by mitosis. The sexual reproductive process unites the gametes to produce a **zygote** that begins the diploid **sporophyte generation** phase.

In lower plants (mosses, liverworts), the sporophyte is small and dependent upon the gametophyte. However, in higher plants (ferns, gymnosperms, angiosperms), the male gametophyte generation is reduced to a tiny pollen tube and three haploid nuclei (called the microgametophyte). The female gametophyte (called the megagametophyte) is a single

multinucleated cell, also called the **embryo sac**. The genotype of the gametophyte or sporophyte influences sexual reproduction in species with self-incompatibility problems. This has implications in the breeding of certain plants; this is discussed further later in this chapter.

5.4.2 Duration of plant growth cycles

The plant breeder should know the lifecycle of the plant to be manipulated. The strategies for breeding are influenced by the duration of the plant growth cycle. Angiosperms (flowering plants) may be classified into four categories based on the duration of their growth cycle (Figure 5.2):

- Annual.** Annual plants (or annuals) complete their lifecycle in one growing season. Examples of such plants include corn, wheat, and sorghum. Annuals may be further categorized into **winter annuals** or **summer annuals**. Winter annuals (e.g. wheat) utilize parts of two seasons. They are planted in the fall (autumn) and undergo a critical physiological inductive change called **vernalization**, which is required for flowering and fruiting in spring. In cultivation, certain non-annuals (e.g. cotton) are produced as though they were annuals.

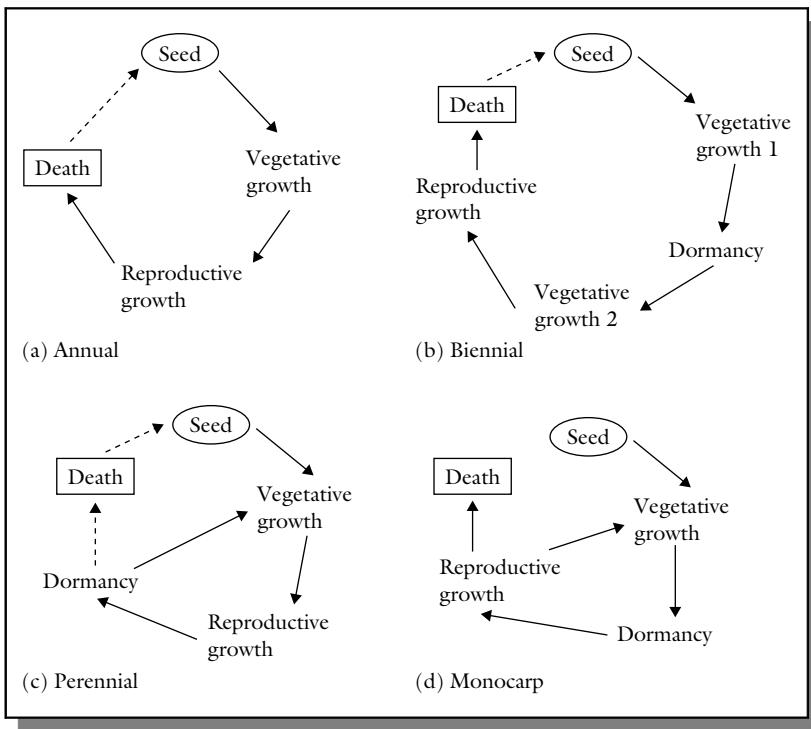


Figure 5.2 Flowering plants have one of four lifecycles – annual, biennial, perennial, and monocarp. Variations occur within each of these categories, partly because of the work of plant breeders.

- (ii) **Biennial.** A biennial completes its lifecycle in two growing seasons. In the first season it produces basal roots and leaves; then it grows a stem, produces flowers and fruits, and dies in the second season. The plant usually requires a special environmental condition or treatment (e.g., vernalization) to be induced to enter the reproductive phase. For example, sugar beet grows vegetatively in the first season. In winter, it becomes vernalized and starts reproductive growth in spring.
- (iii) **Perennial.** Perennials are plants that have the ability to repeat their lifecycles indefinitely by circumventing the death stage. They may be herbaceous, as in species with underground vegetative structures called **rhizomes** (e.g. indiangrass), or aboveground structures called **stolons** (e.g. buffalograss). They may also be woody, as in shrubs, vines (grape), and trees (orange).
- (iv) **Monocarp.** Monocarps are annuals or biennials, but some persist in vegetative development for very long periods (e.g., the so-called “century plant”) before they flower and set seed (e.g.,

bamboo and agave). Once flowering occurs, the plant dies. That is, monocarps are plants that flower only once. Other examples are bromeliads. The top part dies, so that new plants arise from the root system of the old plant.

It should be pointed out that certain plants that may be natural biennials or perennials are cultivated by producers as annuals. For example, sugar beet, a biennial, is commercially produced as an annual for its roots. For breeding purposes it is allowed to bolt to produce flowers for crossing, and subsequently to produce seed.

5.5 Autogamy

Self-pollination or **autogamy** occurs in a wide variety of plant species – vegetables (lettuce, tomatoes, snap beans, endive), legumes (soybean, peas, lima beans) and grasses (barley, wheat, oats). Certain natural mechanisms promote or ensure self-pollination, specifically cleistogamy and chasmogamy, while other mechanisms prevent self-pollination (e.g., self-incompatibility, male sterility).

5.5.1 Mechanisms that promote autogamy

Cleistogamy is the condition in which the flower fails to open. The term is sometimes extended to mean a condition in which the flower opens only after it has been pollinated (as occurs in wheat, barley, lettuce), a condition called **chasmogamy**. Some floral structures, such as those found in legumes, favor self-pollination. Sometimes, the stigma of the flower is closely surrounded by anthers, making it prone to selfing.

Very few species are completely self-pollinated. The level of self-pollination is affected by factors including the nature and amount of insect pollination, air current, and temperature. In certain species, pollen may become sterilized when the temperature dips below freezing. Any flower that opens prior to self-pollination is susceptible to some cross-pollination. A list of predominantly self-pollinated species is presented in Table 5.2.

5.5.2 Mechanisms that prevent autogamy

There are several mechanisms in nature that work to prevent self-pollination in species that otherwise would be self-pollinated. These include self-incompatibility, male sterility and dichogamy.

Self-incompatibility

Self-Incompatibility (or lack of self-fruitfulness) is a condition in which the pollen from a flower is not receptive on the stigma of the same flower, and hence incapable of setting seed. This happens in spite of the fact that both pollen and ovule development are normal and viable. It is caused by a genetically controlled physiological hindrance to self-fertilization. Self-incompatibility is widespread in nature, occurring in families such as Poaceae, Cruciferae, Compositae, and Rosaceae. The incompatibility reaction is genetically conditioned by a locus designated *S*, with multiple

Table 5.2 Examples of predominantly self-pollinated species.

Common name	Scientific name
Barley	<i>Hordeum vulgare</i>
Chickpea	<i>Cicer arietinum</i>
Clover	<i>Trifolium</i> spp.
Common bean	<i>Phaseolus vulgaris</i>
Cotton	<i>Gossypium</i> spp.
Cowpea	<i>Vigna unguiculata</i>
Eggplant	<i>Solanum melongena</i>
Flax	<i>Linum usitatissimum</i>
Jute	<i>Corchorus escularpis</i>
Lettuce	<i>Lactuca sativa</i>
Oat	<i>Avena sativa</i>
Pea	<i>Pisum sativum</i>
Peach	<i>Prunus persica</i>
Peanut	<i>Arachis hypogaea</i>
Rice	<i>Oryza sativa</i>
Sorghum	<i>Sorghum bicolor</i>
Soybean	<i>Glycine max</i>
Tobacco	<i>Nicotiana tabacum</i>
Tomato	<i>Solanum lycopersicum</i>
Wheat	<i>Triticum aestivum</i>

alleles that can number over 100 in some species such as *Trifolium pretense*. However, unlike monoecy and dioecy, all plants produce seed in self-incompatible species.

Male sterility

Male sterility is a condition in plants whereby the anthers or pollen are non-functional. The condition may manifest most commonly as absence of or extreme scarcity of pollen, severe malformation or absence of flowers or stamens, or failure of pollen to dehisce. Just like self-incompatibility, male sterility enforces cross-pollination. Similarly, it can be exploited as a tool to eliminate the need for emasculation for producing hybrid seed. There are three basic kinds of male sterility based on the origin of the abnormality:

- (i) **True male sterility** – This is due to unisexual flowers that lack male sex organs (dioecy and monoecy), or bisexual flowers with abnormal or non-functional microspores (leading to pollen abortion).
- (ii) **Functional male sterility** – The anthers fail to release their contents even though the pollen is fertile.
- (iii) **Induced male sterility** – Plant breeders may use chemicals to induce sterility.

Dichogamy

Dichogamy is the maturing of pistils and stamens of a flower at different times. When this occurs in a self-pollinated species, opportunities for self-pollination are drastically reduced or eliminated altogether, thus making the plant practically cross-pollinated. There are two forms of dichogamy – **protogyny** (stigma is receptive before the anther is mature to release the pollen) and **protandry** (pollen is released from the anther before the female is receptive).



6

Allogamy

Purpose and expected outcomes

There are major world crops that reproduce via allogamy. The breeding methods for autogamy are different from those for allogamy because the mode of reproduction has such profoundly different genetic consequences. After studying this chapter, the student should be able to:

- 1 Discuss the natural mechanisms that favor allogamy.
- 2 Discuss the genetic consequences of allogamy.
- 3 Discuss the implications of allogamy in crop improvement.

6.1 What is allogamy?

Allogamy occurs when fertilization of the flower of a plant is effected by pollen donated by a different plant within the same species. This is synonymous with **cross-pollination, cross-fertilization or outbreeding**, involving the actual fusion of gametes (sperm and ovum). An (incomplete) list of allogamous species is presented in Table 6.1.

6.2 Mechanisms that favor allogamy

Allogamous species depend on agents of pollination, especially wind and insects, and hence tend to produce large amounts of pollen and have large, bright-colored fragrant flowers to attract insects. They commonly have taller stamens than carpels or use other

mechanisms to better ensure the dispersal of pollen to other plants flowers. Other provisions that promote cross-fertilization are mechanisms that control the timing of the receptiveness of the stigma and shedding of pollen and, thereby, prevent autogamy within the same flower. In **protandry**, the anthers release their pollen before the stigma of the same flower is receptive (protandrous flower). In **protogyny**, the stigma is receptive before the pollen is shed from the anthers of the same flower (protogynous flower). Several mechanisms occur in nature by which cross-pollination is ensured, the most effective being dioecy, monoecy, dichogamy, and self-incompatibility. Some mechanisms are stringent in enforcing cross-pollination (e.g., dioecy), while others are less so (e.g., monoecy). These mechanisms are exploited by plant breeders during the controlled pollination phase of their breeding programs, so that

Table 6.1 Examples of predominantly cross-pollinated species.

Common name	Scientific name
Alfalfa	<i>Medicago sativa</i>
Annual ryegrass	<i>Lolium multiflorum</i>
Banana	<i>Musa</i> spp.
Birdsfoot trefoil	<i>Lotus corniculatus</i>
Cabbage	<i>Brassica oleracea</i>
Carrot	<i>Daucus carota</i>
Cassava	<i>Manihot esculentum</i>
Cucumber	<i>Cucumis sativa</i>
Fescue	<i>Festuca</i> spp.
Kentucky bluegrass	<i>Poa pratense</i>
Maize	<i>Zea mays</i>
Muskmelon	<i>Cucumis melo</i>
Onion	<i>Allium</i> spp.
Potato	<i>Solanum tuberosum</i>
Radish	<i>Raphanus sativus</i>
Rye	<i>Secale cereale</i>
Sugarbeet	<i>Beta vulgaris</i>
Sunflower	<i>Helianthus annuus</i>
Sweet potato	<i>Ipomoea batatas</i>
Watermelon	<i>Citrullus lanatus</i>

Though predominantly pollinated, some of these species may have another reproductive mechanism in breeding and crop cultural systems. For example, banana is vegetatively propagated (and not grown from seed), as are cassava and sweet potato; cabbage and maize are produced as hybrids.

only desired pollen sources participate in siring the next plant generation.

6.2.1 Monoecy

Some flowers are complete (possess all the four basic parts) while others are incomplete (are missing one or more of the four basic floral parts). Furthermore, in some species, the sexes are separate. When separate male and female flowers occur on the same plant, the condition is called **monoecy**. Sometimes, the male and female flowers occur in different kinds of inflorescence (different locations, as in corn). Other examples of monoecious plant include most figs, birch, and pine trees. It is easier and more convenient to self-pollinate plants when the sexes occur in the same inflorescence. In terms of seed production, monoecy and dioecy may appear to be inefficient because not all flowers produce seed. Some flowers produce only pollen.

6.2.2 Dioecy

When the sexes occur on different plants (i.e., there are female plants and male plants), the condition is called **dioecy**. Examples of dioecious crop species include date, hops, asparagus, spinach, holly, and hemp. The separation of the sexes means that, by necessity, all seed from dioecious species are hybrid in composition. Where the economic product is the seed or fruit, it is imperative to have female and male plants in the field in an appropriate ratio. In dioecious fruit orchards (e.g., date, persimmons), 3–4 males per 100 females may be adequate. In hops, the commercial product is the female inflorescence. Unfertilized flowers have the highest quality. Consequently, it is not desirable to grow pollinators in the same field when growing hops. Dioecious crops propagated by seed may be improved by mass selection or controlled hybridization.

6.3 Genetic and breeding implications of allogamy

The genotype of the sporophytic generation is highly heterozygous while the genotypes of gametes of a single plant are all different. The genetic structure of a cross-pollinated species is characterized by a high level of heterozygosity. However, this is not to say that at each locus heterozygosity occurs. Especially when the allele frequency of certain genes is high (Chapter 3), a plant may very well be homozygous for that locus. Another source of some homozygosity may be due to occasional selfing in a plant. Unlike allogamous species in which formation of new gene combinations are discouraged, cross-pollinated species share a wide gene pool from which new combinations are created to form the next generation.

It is instructive to state that in autogamous crops in principle the whole genotype is transmitted through the generations (i.e., they are “immortal”). Homozygous plants reproduce genetically identically. Consequently, the unit of selection in a mixture of homozygous lines is genotype. In contrast, in allogamous crops the unit of selection is the single gene. The gene in this case is “immortal.” Genotypes perish (lose their identity) at each round of sexual reproduction. The only way the genotype can become immortal and be the unit of selection in allogamous crops is when they are clonally propagated, as is the case in potato.

Allogamous species may undergo self-fertilization to a varying extent. In that case the progeny usually suffers from **inbreeding depression**. deleterious recessive alleles that were suppressed because of heterozygous advantage have opportunities to become homozygous, and therefore become expressed. However, such depression is reversed upon cross-pollination. **Hybrid vigor** (the increase in vigor of the hybrid over its partially homozygous and distinct parents) is exploited in hybrid seed production (Chapter 18). In addition to hybrid breeding, population-based improvement methods (e.g., mass selection, recurrent selection, and synthetic cultivars) are common methods of breeding cross-pollinated species.

6.4 Inbreeding depression

As previously stated, inbreeding or crossing closely related parents results in reduced fitness or vigor of individuals in the progenitor population, a condition called inbreeding depression. Reduction in fitness usually manifests itself as a reduction in vigor, fertility, and productivity, and is seen as lower biomass per plant, lower fecundity, malformation of organs and lower germination of seeds. The effect of inbreeding is more severe in the early generations (5–8) than in later generations. Just like heterosis, inbreeding depression is not uniformly manifested in plants. Plants including onions, sunflower, cucurbits, maize and rye are rather tolerant of inbreeding with low or no inbreeding depression. On the other hand, crops such as alfalfa and carrot are highly intolerant of inbreeding.

6.4.1 The concept of genetic load

Genetic load (or genetic burden) may be defined as the decrease in fitness of the average individual in a population due to the presence of deleterious genes or genotypes in the gene pool. In other words, it is the reduction in selective value for a population compared to what it would otherwise have if all the individuals had the most favored genotype. Statistically, its value ranges between zero (no load) and one. It is generally believed that most species carry a genetic load of 3–5 recessive lethal genes. The genes are mostly hidden (enjoy heterozygous advantage). Inbreeding usually causes the genetic load to increase. Genetic load has three components – mutational load (due to harmful mutations),

segregational load (due to heterozygous advantage), and substitutional or frequency-dependent load (occurs during transient polymorphism; it arises in a population in which natural selection is acting to substitute one allele for another). Genetic load generally lowers the viability of a population.

Because of selection, the frequency of deleterious recessive alleles in a population is expected to decrease rapidly with higher levels of inbreeding. Eventually, these alleles may be lost from the population, a process sometimes referred to as purging populations of their genetic load. Populations that have experienced long periods of inbreeding are expected to show less inbreeding consequences.

6.5 Hybrid vigor

Hybrid vigor or **heterosis** is opposite and complementary to inbreeding depression (reduction in fitness as a direct result of inbreeding). In theory, the heterosis observed after crossing is expected to be equal to the depression upon inbreeding, considering a large number of crosses between lines derived from a single base population. In practice, plant breeders are interested in heterosis expressed by specific crosses between selected parents, or between populations that have no known common recent origin. Furthermore, because heterosis is subject to the interactions between genotype and environment, it is desirable to describe the heterosis of a particular hybrid line for a specific trait at a specific location or under specified environmental conditions.

Hybrid vigor may be defined as the increase in size, vigor, fertility, and overall productivity of a hybrid plant, F_1 , over the mid-parent value (average performance of the two parents P_1 and P_2). It is calculated as the difference between the cross-bred and inbred means.

$$\text{Hybrid vigor} = \{[F_1 - (P_1 + P_2)/2]/[(P_1 + P_2)/2]\}$$

The estimate is usually calculated as a percentage (i.e., $\times 100$).

The synonymous term, **heterosis**, was coined by G.H. Shull. Heterosis is of little commercial value (and hencevalue to the farmer) if a hybrid will only exceed the mid-parent in performance. Hence, the *practical* definition of heterosis is hybrid vigor that greatly exceeds the better or higher parent in a cross. Such advantageous hybrid vigor is observed, in

particular, when breeders cross parents that are genetically diverse. Heterosis occurs when two inbred lines or outbred species are crossed.

In theory, heterosis may be “positive” or “negative”. This is largely an artificial distinction. Positive heterosis is generally desired for traits like yield, while negative heterosis is desired for traits such as early maturity. Three kinds of heterosis may be distinguished as – mid-parent, standard variety, and better parent (also called heterobeltiosis). Standard variety (or check) heterosis is measured by comparing the hybrid to existing high yield commercial variety. Considering the fact that breeders aim to develop cultivars that excel in performance to existing commercial ones, standard variety heterosis is perhaps most desirable to breeders.

Heterosis, though widespread in the plant kingdom, is not uniformly manifested in all species and for all traits. It is manifested at a higher intensity in traits that have fitness value, and also more frequently and at higher levels among cross-pollinated species than self-pollinated species. All breeding methods that are preceded by crossing make use of heterosis to some extent. However, it is only in hybrid cultivar breeding and the breeding of clonally propagated varieties that the breeder has the opportunity to exploit the phenomenon to full advantage.

Hybrids may have dramatically increased yields compared to open-pollinated cultivars. By the early 1930s (before extensive use of hybrids), maize yield in the United States averaged 1250 kg/ha. By the early 1970s (following the adoption of hybrids), maize yields quadrupled to 4850 kg/ha. The contribution of hybrids (genotype) to this increase was estimated at about 60%, the remainder being attributed to production practices.

6.5.1 Genetic basis of heterosis

Three schools of thought have been advanced to explain the genetic basis for why fitness lost on inbreeding tends to be restored upon crossing. The two most commonly known are the **dominance theory**, first proposed by C.G. Davenport in 1908 and later by I.M. Lerner, and the **overdominance theory**, first proposed by G. H. Shull in 1908 and later by K. Mather and J.L. Jinks. A third theory, the mechanism of epistasis (non-allelic gene interactions), has also been proposed by researchers (such as A.C. Fasoulas and R.W. Allard in 1962). Any viable theory should

account for both inbreeding depression in cross-pollinated species upon selfing and increased vigor in F_1 , upon hybridization. It should be pointed out that the proposed mechanisms do not occur in exclusion to one another but indeed could operate simultaneously, each in different genes. Further, even though the dominance theory is the most favored by most scientists, none of the theories is completely satisfactory.

Dominance theory

The dominance theory assumes that vigor in plants is conditioned by dominant (functional) alleles, recessive alleles being deleterious or neutral in effect, mostly representing loss-of-function versions of the original dominant gene. It follows then that a genotype with more dominant alleles will be more vigorous than one with few dominant alleles. Consequently, inbreeding parents that are homozygous dominant or heterozygous at most loci will be vigorous but upon inbreeding heterozygous loci may result in progeny that is homozygous for recessive non-functional alleles at several or many loci, resulting in inbreeding depression. If such inbreeding is done on two parents that are of distinct origin, chances are low that they will carry deleterious alleles for the same loci. Therefore, crossing two such largely homozygous parents with complementary dominant and recessive alleles will concentrate more favorable alleles in the hybrid than either inbred parent. In practice, linkage and the large number of genes to be taken care of prevent the breeder from developing inbred lines that contain all dominant alleles in homozygous state. Inbreeding depression occurs upon selfing because the deleterious recessive alleles that are protected in the heterozygous condition (heterozygous advantage) become homozygous and are expressed. In corn, inbred lines have been developed with a limited number and limited deleteriousness of homozygous recessive alleles, resulting in only limited inbreeding depression. These inbred lines are sufficiently fit to produce enough seeds to serve as parents for hybrid cultivar seed production.

To illustrate this theory, assume a quantitative trait like seed yield is conditioned by four loci. Assume that each allele in the dominant homozygous or heterozygous state contributes two units to the phenotype, while a recessive homozygous genotype contributes one unit. A cross between two inbred parents produces the following outcome:

	P_1 $(AAbbCCdd)$	x	
Phenotypic value	$2 + 1 + 2 + 1 = 6$	↓	$1 + 2 + 1 + 2 = 6$
F_1		$(A\alpha BbCcDd)$ $2 + 2 + 2 + 2 = 8$	

because the homozygous and heterozygous dominant state will both contribute two units to the phenotype. The result is that the F_1 would be more productive than either parent.

D.L. Falconer developed a mathematical expression for the relationship between the parents in a cross that leads to heterosis as follows:

$$HF_1 = \Sigma dy^2$$

where HF_1 is the deviation of the hybrid from the mid-parent value, d is the degree of dominance, and y

gene (e.g., A , α) are contrasting but each has a different favorable effect in the plant. In this view allele α is not supposed to have a loss of function. Consequently, a heterozygous locus would have greater positive effect than either homozygous locus and, by extrapolation, a genotype with more heterozygous loci would be more vigorous than one with less heterozygous loci.

To illustrate this phenomenon, consider a quantitative trait conditioned by four loci. Assume that recessive, heterozygote, and homozygote dominants contribute 1, 2, and $1\frac{1}{2}$ units to the phenotypic value, respectively:

	P_1 $(AAbbCCdd)$	x	
Phenotypic value	$1\frac{1}{2} + 1 + 1\frac{1}{2} + 1 = 5$	↓	$(\alpha\alpha BBccDD)$ $1 + 1\frac{1}{2} + 1 + 1\frac{1}{2} = 5$
		$(A\alpha BbCcDd)$ $2 + 2 + 2 + 2 = 8$	

is the difference in gene frequency in the parents of the cross. From the expression, maximum mid-parent heterosis (HF_1) will occur when the values of the two factors (d , y) are each unity. That is, the populations to be crossed are fixed for opposite alleles ($y = 1.0$) and there is complete dominance ($d = 1.0$).

Overdominance theory

The phenomenon of a heterozygote being superior to the best performing homozygote is called **overdominance** (i.e., heterozygosity *per se* is assumed to be responsible for heterosis). A possible explanation for this could be the fact that genes normally have pleiotropic effects and, thereby, contribute simultaneously to many measurable traits of the plant. The overdominance theory assumes that the alleles of a

Heterozygosity leads to the highest trait values of the three genotypes.

Where the dominance theory applies, heterosis, theoretically, can be fixed in a pure line; however, where overdominance applies, this cannot occur. Of course, both theories are not exclusive. Some types of gene may contribute to heterosis because of the dominance effect, others because of the overdominance effect.

6.5.3 Factors to consider in using heterosis in breeding

Springer and Stupar (2007) summarized four factors to note when considering the application of heterosis in crop improvement:

- (i) The magnitude of heterosis is variable among species. The effects of heterosis are stronger and more ubiquitous in corn (allogamous species) than in, say, *Arabidopsis* (autogamous species).
- (ii) The level of heterosis for specific traits varies and is not correlated in different hybrids of the same species. This indicates that the phenomenon of heterosis is not conditioned by the action of a single locus, nor does it simply represent the overall extent of heterozygosity between parents.
- (iii) Generally, heterosis increases as the genetic distance between the parental inbreds increases. However, there is a threshold that when genetic distance between the parents is exceeded, heterosis decreases. Whereas this appears to suggest a relationship between genetic diversity and heterosis, this relationship is not strong enough to make it a predictive tool.
- (iv) The allelic variation that produces heterosis does not represent the totality of variation that occurs. Not all allelic variants in a species' population will be fixed in inbred lines because variants with strong deleterious phenotypes will be selected against by breeders. Consequently, the range of allelic variation in inbred lines that can contribute to heterosis is limited to only the variation with beneficial effects for a specific trait or that which has limited deleterious effects. In other words, not all allelic variation between parental pairs contributes to heterosis. Some allelic variation will not be fixed, that is, when the homozygous state of a certain allele is deleterious to the genotype.



9

Variation: types, origin and scale

Purpose and expected outcomes

Biological variation is a fact of nature. No two plants are exactly alike. Plant breeders routinely deal with variability in one shape or form. It is indispensable to plant breeding, and hence breeders assemble or create it as a critical first step in a breeding program. Then, they have to discriminate among the variability, evaluate and compare superior genotypes, and increase and distribute the most desirable to consumers. After completing this chapter, the student should be able to:

- 1 Discuss the types of variation.
- 2 Discuss the origins of genetic variation.
- 3 Discuss the scale of genetic variation.
- 4 Distinguish between qualitative and quantitative variation.
- 5 Discuss the rules of plant classification.

9.1 Classifying plants

Plant taxonomy is the science of classifying and naming plants. Organisms are classified into five major groups (kingdoms) – Plantae, Animalia, Fungi, Protista, and Monera (Table 9.1). Plant breeders are most directly concerned about Kingdom Plantae, the plant kingdom. However, one of the major objectives of plant breeding is to breed for resistance of the host to diseases and economic destruction caused by organisms in the other four kingdoms that adversely

impact plants. Plant breeding depends on plant variation or diversity for success. It is critical that the appropriate plant material is acquired for a breeding program. An international scientific body sets the rules for naming plants. Standardizing the naming of plants eliminates the confusion from the numerous culture-based names of plants. For example, corn in the United States is called maize in Europe, not to mention the thousands of other names worldwide.

The **binomial nomenclature** was developed by Carolus Linnaeus and entails assigning two names

Table 9.1 The five kingdoms of organisms as described by Whitaker.

1. **Monera** (have prokaryotic cells)
 - Bacteria
2. **Protista** (have eukaryotic cells)
 - Algae
 - Slime molds
 - Flagellate fungi
 - Protozoa
 - Sponges
3. **Fungi** (absorb food in solution)
 - True fungi
4. **Plantae** (produce own food by the process of photosynthesis)
 - Bryophytes
 - Vascular plants
5. **Animalia** (ingest their food)
 - Multicellular animals

based on the genus and species, the two bottom taxa in taxonomic hierarchy (Figure 9.1). It is important for the reader to understand that plant breeding by conventional tools alone is possible primarily at the species level. Crosses are possible within species and occasionally between species (but often problematic). However, plant breeding incorporating molecular tools allows gene transfer from any taxonomic

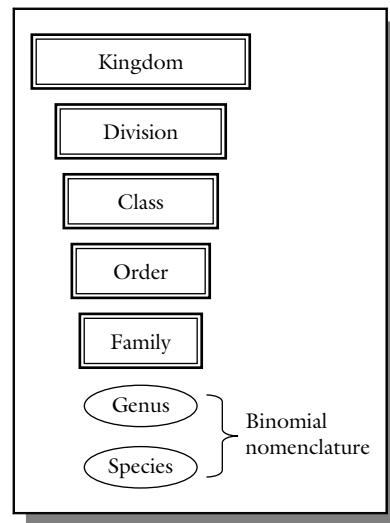


Figure 9.1 Taxonomic hierarchy of plants. Plant breeders routinely cross plants without problem within a species. Crosses between species are problematic and often impossible between genera and beyond.

level to another. It is important to emphasize that such a transfer is not routine and has its challenges.

The kingdom Plantae comprises **vascular plants** (plants that contain conducting vessels – xylem and phloem) and **non-vascular plants** (Table 9.2) Vascular plants may be seeded or seedless. Furthermore,

Table 9.2 Divisions in the kingdom Plantae.

	Division	Common Name
Bryophytes	Hepaticophyta Anthocerotophyta Bryophyta	Liverworts Hornworts Mosses
Vascular plants		
Seedless	Psilotophyta Lycophyta Sphenophyta Pterophyta	Whisk ferns Club mosses Horsetails Ferns
Seeded	Pinophyta Subdivision: Cycadidae Subdivision: Pinidae Class: Ginkgoatae Class: Pinatae Subdivision: Gnetales Magnoliophyta Class: Liliopsida Class: Magnoliopsida	Gymnosperms Cycads <i>Ginko</i> Conifers <i>Gnetum</i> Flowering plants Monocots Dicots

Table 9.3 Important field crop families in the division Magnoliophyta (flowering plants).**Monocots****1 Poaceae (grass family)**

In terms of numbers, the grass family is the largest of flowering plants. It is also the most widely distributed.

Examples of species: wheat, barley, oats, rice, corn, fescues, and bluegrass.

2 Aracaceae (palm family)

The palm family is tropical and subtropical in adaptation.

Examples of species: oil palm (*Elaeis guineensis*), coconut palm (*Cocos nucifera*).

3 Amaryllidaceae (amaryllis family)

Plants with tunicate bulbs characterize this family.

Examples of species: onion, garlic, and chives.

Dicots**1 Brassicaceae (mustard family)**

The mustard family is noted for its pungent herbs.

Examples of species: cabbage, radish, cauliflower, turnip and broccoli.

2 Fabaceae (legume family)

The legume family is characterized by flowers that may be regular or irregular. The species in this family are an important source of protein for humans and livestock.

Examples of species: dry beans, mung bean, cowpea, pea, peanut, soybean, and clover.

3 Solanaceae (night shade family)

This family is noted for the poisonous alkaloids many of them produce (e.g. belladonna, nicotine, atropine, and solanine).

Examples of species: tobacco, potato, tomato, pepper, and eggplant.

4 Euphorbiaceae (spurge family)

Members of the spurge family produce milky latex and include a number of poisonous species.

Examples of species: cassava (*Manihot esculenta*) and castor bean.

5 Asteraceae (sunflower family)

The sunflower family has the second largest number of flowering plant species.

Example of species: sunflower and lettuce.

6 Apiaceae (carrot family)

Plants in this family usually produce flowers that are arranged in umbels.

Examples of species: carrot, parsley, and celery.

7 Cucurbitaceae (pumpkin family)

The pumpkin or gourd family is characterized by prostrate or climbing herbaceous vines with tendrils and large, fleshy fruits containing numerous seeds.

Examples of species: pumpkin, melon, watermelon, and cucumber.

seeded plants may be **gymnosperms** (have naked seed) or **angiosperms** (have seed borne in a fruit). Flowering plants may have seed with one cotyledon, called **monocots** (these include grasses such as wheat, barley and rice), or seed with two cotyledons, called **dicots** (these include legumes such as soybean, pea, and peanut) (Table 9.3). The strategies for breeding flowering species are different from those for non-flowering species. Flowering species can be genetically manipulated through the sexual process (sexually reproducing) by crossing, whereas non-flowering species (asexually reproducing)

cannot. Furthermore, even within flowering plants, the method for breeding differs according to the mode of pollination – self-pollination or cross-pollination.

9.3 Operational classification systems

Crop plants may be classified for specific purposes, for example, according to seasonal growth, kinds of stem, growth form, and economic part or agronomic use.

- (a) **Seasonal growth cycle.** Plants may be classified according to the duration of their lifecycle (i.e. from seed, to seedling, to flowering, to fruiting, to death, and back to seed). On this basis, crop plants may be classified as **annual**, **biennial**, **perennial**, or **monocarp**, as previously discussed in Chapter 4.
- (b) **Stem type.** Certain plants have non-woody stems, existing primarily in vegetative form (e.g., onion, corn, or sugar beet) and are called **herbs** (or herbaceous plants). **Shrubs** are plants with multiple stems that arise from the ground level (e.g., dogwood, azalea, kalmia) while **trees** (e.g. apple, citrus, palms) have one main trunk or central axis.
- (c) **Common stem growth form.** Certain plants can stand upright without artificial support; others cannot. Based on this characteristic, plants may be classified into groups. The common groups are.
- **Erect.** Erect plants can stand upright without physical support, growing at about a 90° angle to the ground. This feature is needed for mechanization of certain crops during production. Plant breeders develop erect (bush) forms of non-erect (pole) cultivars for this purpose. There are both pole and bush cultivars of crops such as bean (*Phaseolus vulgaris* L.) in cultivation.
 - **Decumbent.** Plants with decumbent stem growth form, such as peanuts (*Arachis hypogaea*), are extremely inclined with raised tips.
 - **Creeping (or repent).** Plants in this category, such as strawberry (*Fragaria* spp.), have stems that grow horizontally on the ground.
 - **Climbing.** Climbers are plants with modified vegetative parts (stems or leaves) that enable them to wrap around a nearby physical support, so they do not have to creep on the ground. Examples are yam (*Dioscorea* spp.) and ivy.
 - **Despitose (bunch or tufted).** Grass species, such as buffalo grass, have a creeping form whereas others, such as tall fescue, have a bunch form and hence do not spread by horizontal growing stems.
- (d) **Agronomic use.** Crop plants may be classified according to agronomic use as follows:
- **Cereals** – These are grasses such as wheat, barley, and oats that are grown for their edible seed.
 - **Pulses** – Legumes grown for their edible seed (e.g. peas, beans).
 - **Grains** – Crop plants grown for their edible dry seed (e.g. corn, soybean, cereals).
 - **Small grains** – Grain crops with small seed (e.g. wheat, oats, barley).
 - **Forage** – Plants grown for their vegetative matter that is harvested and used fresh or preserved as animal feed (e.g. alfalfa, red clover).
 - **Roots** – Crops grown for their edible modified (swollen) roots (e.g. sweet potato, cassava).
 - **Tubers** – Crops grown for their edible modified (swollen) stem (e.g. Irish potato, yam).
 - **Oil crops** – Plants grown for their oil content (e.g. soybean, peanut, sunflower, oil palm).
 - **Fiber crops** – Crop plant grown for use in fiber production (e.g. jute, flax, cotton).
 - **Sugar crops** – Crops grown for use in making sugar (e.g. sugar cane, sugar beet).
 - **Green manure crops** – Crop plant grown and plowed under the soil while still young and green, for the purpose of improving soil fertility (e.g. many leguminous species).
 - **Cover crops** – Crops grown between regular cropping cycles, for the purpose of protecting the soil from erosion and other adverse weather factors (e.g. many annuals).
 - **Hay** – Grasses or legume plants that are grown, harvested and cured for feeding animals (e.g. alfalfa, buffalo grass).
- There are other operational classifications used by plant scientists.

(e) **Adaptation.** Plants may be classified on the basis of temperature adaptation as either cool season or warm season plants.

- **Cool season or temperate plants.** These plants, such as wheat, sugar beet, and tall fescue prefer a monthly temperature of between 15 and 18°C (59–64°F) for growth and development.
- **Warm season or tropical plants.** These plants, such as corn, sorghum, and buffalo grass require warm temperatures of between 18 and 27°C (64–80°F) during the growing season.

Whereas some of the operational classifications are applicable, horticultural plants have additional classification systems. These include the following:

- **Fruit type**
 - 1 **Temperate fruits** (e.g., apple, peach) versus **tropical fruits** (e.g., orange, coconut).
 - 2 **Fruit trees** – have fruits borne on trees (e.g., apple, pear).
 - 3 **Small fruits** – generally woody perennial dicots (e.g., strawberry, blackberry).
 - 4 **Bramble fruits** – non-tree fruits that need physical support (e.g., raspberry).
- **Flowering** (sunflower, pansy) versus **foliage** (non-flowering – coleus, sansevieria) plants.
- **Bedding plants** – annual plants grown in beds (e.g., zinnia, pansy, petunia).
- **Deciduous trees** (shed leaves seasonally) versus **evergreen plants** (no leaf shedding).

9.4 Types of variation among plants

As previously indicated, the phenotype (the observed trait) is the product of the genotype and the environment ($P = G + E$). The phenotype may be altered by altering G, E, or both. There are two fundamental sources of variation in phenotype: genotype and the environment – and hence two kinds of variation – **genetic** and **environmental**. Later in the book an additional source of variation, $G \times E$ (interaction of the genotype and the environment), will be introduced.

9.4.1 Environmental variation

When individuals from a clonal population (i.e., identical genotype) are grown in the field, the plants will exhibit differences in the expression of some traits because of non-uniform environment. The field is often heterogeneous with respect to plant growth factors – nutrients, moisture, light, and temperature. Some fields are more heterogeneous than others. Sometimes, non-growth factors may occur in the environment and impose different intensities of environmental stress on plants. For example, disease and pest agents may not uniformly infect plants in the field. Similarly, plants that occur in more favorable parts of the field or are impacted to a lesser degree by an adverse environmental factor would perform better than disadvantaged plants. That is, even clones may perform differently under different environments and inferior genotypes can outperform superior genotypes under uneven environmental conditions. If a breeder selects an inferior genotype by mistake, the progress of the breeding program will be slowed. Consequently, plant breeders use statistical tools and other selection aids to help in reducing the chance of advancing inferior genotypes, and thereby make rapid progress in the breeding program.

As previously noted, environmental variation is not heritable. However, it can impact heritable variation. Plant breeders want to be able to select a plant on the basis of its nature (genetics) not nurture (growth environment). To this end, evaluations of breeding material are conducted in a uniform environment as much as possible. Furthermore, the selection environment is often similar to the one in which the crop is commercially produced.

9.4.2 Genetic variability

Variability that can be attributed to genes that encode specific traits and can be transmitted from one generation to the next is described as **genetic** or **heritable variation**. Because genes are expressed in an environment, the degree of expression of a heritable trait is impacted by its environment, some more so than others (Figure 4.1). Heritable variability is indispensable to plant breeding. As previously noted, breeders seek to change the phenotype (trait) permanently and heritably by changing the genotype (genes) that encode it. Heritable variability is consistently expressed generation after generation. For example, a purple-flowered genotype will always produce purple flowers. However, a mutation can permanently alter an original expression. For example, a purple-flowered plant may be altered by mutation to become a white-flowered plant.

Genetic variation can be detected at the molecular as well as the gross morphological level. The availability of biotechnological tools (e.g., DNA markers) allows plant breeders to assess genetic diversity of their materials at the molecular level. Some genetic variation is manifested as visible variation in morphological traits (e.g., height, color, size), while compositional or chemical traits (e.g., protein content, sugar content of a plant part) require various tests or devices for evaluating them.

9.5 Origins of genetic variability

There are three ways in which genetic or heritable variability originates in nature – gene recombination, modifications in chromosome number, and mutations. The significant fact to note is that, rather than wait for them to occur naturally, plant breeders use a variety of techniques and methods to manipulate and make these three phenomena more and more targeted, as they generate genetic variation for their breeding programs. With advances in science and technology (e.g., gene transfer, somaclonal variation), new sources of genetic variability have become available to the plant breeder. Variability generated from these sources is, however, so far limited.

9.5.1 Genetic recombination

Genetic recombination applies only to sexually reproducing species and represents the primary source

of variability for plant breeders in those species. As previously described, genetic recombination occurs via the cellular process of meiosis. This phenomenon is responsible for the creation of non-parental types in the progeny of a cross, through the physical exchange of parts of homologous chromosomes (by breakage-fusion). The cytological evidence of this event is the characteristic crossing (X-configuration or **chiasma**) of the adjacent homologous chromosome strands, as described in Chapter 3, allowing genes that were transmitted together (non-independent assortment) in the previous generation to become independent. Consequently, sexual reproduction brings about gene reshuffling and generation of new genetic combinations (recombinants). Unlike mutations that cause changes in genes themselves in order to generate variability, recombination generates variability by assembling new combinations of genes from different parents. In doing this, some gene associations are broken.

Consider a cross between two parents of contrasting genotypes $AAbb$ and $aaBB$. A cross between them will produce an F_1 of genotype $AaBb$. In the F_2 segregating population, and according to Mendel's law, the gametes (AB , Ab , aB , and ab) will combine to generate variability, some of which will be old (like the parents – parental), while others will be new (unlike the parents – recombinants) (Figure 9.2). The

$P_1 = AAbb \times aaBB = P_2$	
	↓
$F_1 = AaBb$	
	↓
F_2	
	AB aB Ab ab
AB	$AABB$ $AaBB$ $AABb$ $AaBb$
	$2^n = 4$ gametes
	$3^n = 9$ genotypes
	$2^n = 4$ different phenotypes

Figure 9.2 Genetic recombination results in the production of recombinants in the segregating population. This phenomenon is a primary source of variability in breeding flowering species. The larger the number of genes (n) the greater the amount of variability that can be generated from crossing.

larger the number of pairs of allelic genes by which the parents differ, the greater the new variability that will be generated. Representing the number of different allelic pairs by n , the number of gametes produced is 2^n , and the number of genotypes produced in the F_2 following random mating is 3^n with 2^n phenotypes (assuming complete dominance). In this example, two new homozygous genotypes ($aabb$, $AABB$) are obtained.

It should be pointed out that recombination only includes genes that are already present in the parents. Consequently, if there is no genetic linkage, the new gene recombination can be predicted. Where linkage is present, knowledge of the distance between gene loci on the chromosomes is needed for estimating their frequencies. As previously discussed in Chapter 3, additional variability for recombination may be observed where intra-allelic and inter-allelic interactions (epistasis) occur. This phenomenon results in new traits which were not found in the parents. Another source of genetic variability is the phenomenon of **gene transgression**, which causes some individuals in a segregating population from a cross to express the trait of interest outside the boundaries of the parents (e.g., taller than the taller parent, or shorter than the shorter parent). These new genotypes are called **transgressive segregants**. The discussion so far has assumed diploidy in the parents. However, in species of higher ploidy levels (e.g., tetraploid, hexaploid), it is not difficult to see how additional genetic variability could result where allelic interactions occur.

One of the tools of plant breeding is hybridization (crossing of divergent parents), whereby breeders selectively mate plants to allow their genomes to be reshuffled into new combinations to generate variability in which selection can be practiced. By carefully selecting the parents to be mated, the breeder has some control over the nature of the genetic variability to be generated. Breeding methods that include repeated hybridization (e.g., reciprocal selection, recurrent selection) offer more opportunities for recombination to occur.

The speed and efficiency with which a breeder can identify (by selecting among hybrids and their progeny) desirable combinations, is contingent upon the number of genes and linkage relationships that are involved. Because linkage is likely to exist, the plant breeder is more likely to make rapid progress with recombination by selecting plant genotypes with

high chiasma frequency (albeit unconsciously). It follows then that the cultivar developed with the desired recombination would also have higher chiasma frequency than the parents used in the breeding program.

9.5.2 Ploidy modifications

New variability may arise naturally through modifications in chromosome number as a result of hybridization (between unidentical genotypes), or abnormalities in the nuclear division processes (spindle malfunction). Failure of the spindle mechanism, during karyokinesis or even prior to that, can lead to errors in chromosome numbers transmitted to cells, such as **polyploidy** (individuals with multiples of the basic set of chromosomes for the species in their cells) (Figure 9.3). Sometimes, instead of variations involving complete sets of chromosomes, plants may be produced with multiples of only certain chromosomes or deficiencies of others (called **aneuploidy**). Sometimes, plants are produced with half the number of chromosomes in the somatic cells (called haploids). Like genetic recombination, plant breeders are able to induce various kinds of chromosome modification

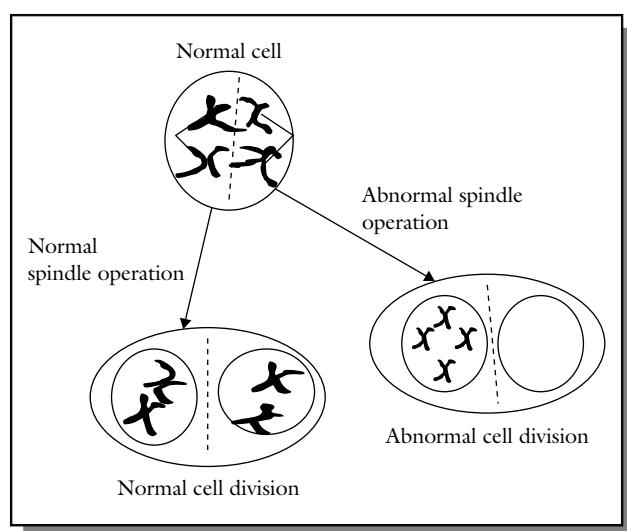


Figure 9.3 Failure of the genetic spindle mechanism may occur naturally or be artificially induced by plant breeders (using colchicine), resulting in cell division products that inherit abnormal chromosome numbers. Plant breeders deliberately manipulate the ploidy of cells to create polyploids.

to generate variability for breeding. The subject is discussed in detail in Chapter 24.

9.5.3 Mutation

Mutation is the ultimate source of biological variation. Mutations are important in biological evolution as sources of heritable variation. They arise spontaneously in nature as a result of errors in cellular processes such as DNA replication (or duplication) and by **chromosomal aberrations** (deletion, duplication, inversion, translocation). The molecular basis of mutation may be described by mechanisms such as modification of the structure of DNA or a component base of DNA, substitution of one base for a different base, deletion or addition of one base in one DNA strand, deletion or addition in one or more base pairs in both DNA strands, and inversion of a sequence of nucleotide base pairs within the DNA molecule. These mechanisms are discussed further in Chapter 23 on mutation breeding.

Mutations may also be induced by plant breeders using agents such as irradiation and chemicals. Many useful mutations have been found in nature or induced by plant breeders (e.g., dwarfs, nutritional quality genes). However, many mutations are deleterious to their carriers and are hence selected against in nature or by plant breeding. From the point of view of the breeder, mutations may be useful, deleterious, or neutral. Neutral mutations are neither advantageous nor disadvantageous to individuals in which they occur. They persist in the population in the heterozygous state as recessive alleles and become expressed only when in the homozygous state, following an event such as selfing.

9.5.4 Transposable elements

The phenomenon of **transposable elements** (genes with the capacity to relocate within the genome), creates new variability. **Transposable genetic elements** (transposable elements, transposons, or “jumping genes”) are known to be nearly universal in occurrence. These mobile genetic units relocate within the genome by the process called **transposition**. The presence of transposable elements indicates that genetic information is not fixed within the genome of an organism.

9.7 Scale of variability

As previously indicated biological variation can be enormous and overwhelming to the user. Consequently, there is a need to classify it for effective and efficient use. Some variability can be readily categorized by counting and placing into distinct non-overlapping groups; this is said to be discrete or **qualitative variation**. Traits that exhibit this kind of variation are called **qualitative traits**. Other kinds of variability occur on a continuum and cannot be placed into discrete groups by counting. There are intermediates between the extreme expressions of such traits. They are best categorized by measuring or weighing and are described as exhibiting continuous or **quantitative variation**. Traits that exhibit this kind of variation are called **quantitative traits**.

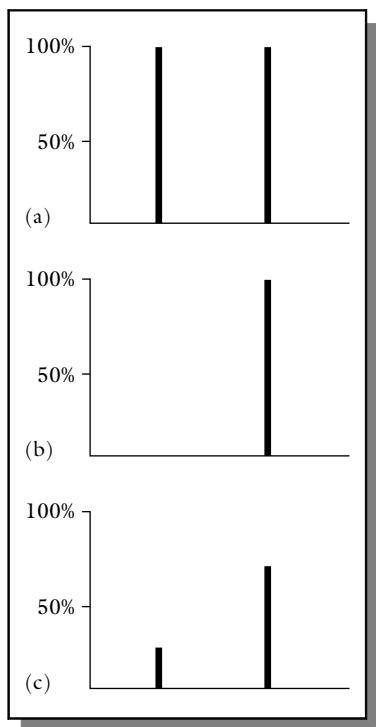


Figure 9.5 Qualitative variation produces discrete measurements that can be placed into distinct categories.

However, there are some plant traits that may be classified either way. Sometimes, for convenience, a quantitative trait may be classified as though it were qualitative. For example, an agronomic trait such as

earliness of plant maturity is quantitative in nature. However, it is possible to categorize cultivars into maturity classes (e.g., in soybean, maturity classes range from 000 (very early) to VIII (very late)). Plant height can be treated in similar fashion, so too is seed coat color (that may be expressed as shades of a particular color).

9.7.1 Qualitative variation

Qualitative variation is easy to classify, study, and utilize in breeding. It is simply inherited (controlled by one or a few genes) and amenable to Mendelian analysis (Figure 9.5). Examples of qualitative traits include diseases, seed characteristics, and compositional traits. Because they are amenable to Mendelian analysis, the chi square statistical procedure may be used to determine the inheritance of qualitative genes. The success of gene transfer using molecular technology so far has involved the transfer of single genes (or a few at best), such as the *Bt* and *Ht* (herbicide tolerant) products.

Breeding qualitative traits

Breeding qualitative traits is relatively straight forward. They are readily identified and selected. Breeding recessive traits is a little different from breeding dominant traits (Figure 9.6). It is important to have a large segregating population, especially if several loci are segregating, to increase the chance of finding the

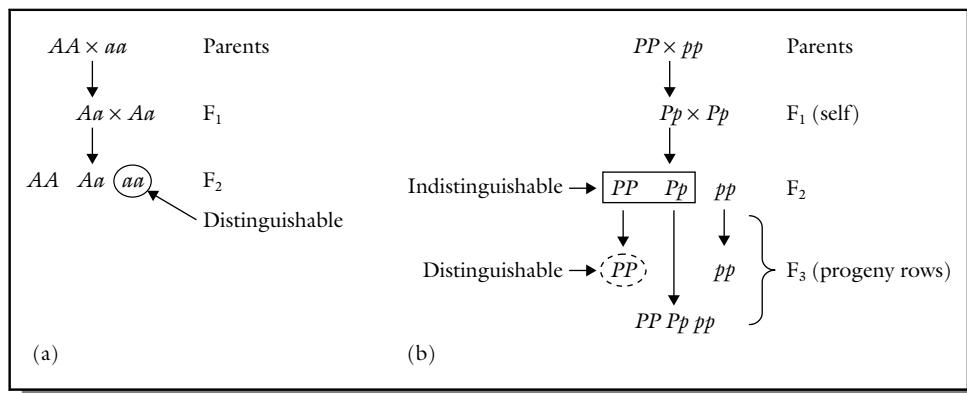


Figure 9.6 Breeding a qualitative trait conditioned by a recessive gene. The desired - *aa* genotype - can be observed and selected in the *F*₂. Breeding a qualitative trait conditioned by a dominant gene. The desired trait cannot be distinguished in the *F*₂, requiring another generation (progeny row) to distinguish between the dominant phenotypes.

desired homozygous recessive genotypes. For example, if two loci are segregating, a cross between $AA \times aa$ would produce 25% homozygous recessive individuals in the F_2 ($1AA:2Aa:1aa$). It is important to note that the desired genotype can be isolated from the F_2 without any further evaluation. In case of a dominant locus, (e.g., the cross $PP \times pp$), 25% of the F_2 will be homozygous recessive, whereas 75% would be of the heterozygous dominant phenotype (of which only 25% would be homozygous dominant). The breeder needs to advance the material one more generation to identify individuals that are homozygous dominant.

9.7.2 Quantitative variation

Most traits encountered in plant breeding are quantitatively inherited. A principal distinguishing feature of this variation is that the trait, whether controlled by few or many genes is influenced significantly by environmental variability, thus resulting in phenotypes that cannot be readily placed in distinct categories like the case for qualitative traits. Variation in quantitative trait expression is hence without natural discontinuities. Traits that exhibit continuous variations are also called **metric traits**. Any attempt to classify such traits into distinct groups is only arbitrary. For example,

height is a quantitative trait. If plants are grouped into tall versus short plants, one could find relatively tall plants in the short group and similarly short plants in the tall group (Figure 9.7).

Quantitative traits are conditioned by many to numerous genes (**polygenic inheritance**) with effects that are too small to be individually distinguished. They are sometimes called **minor genes**. Quantitative trait expression is very significantly modified by the variation in environmental factors to which plants in the population are subjected. Continuous variation is caused by environmental variation and genetic variation due to the simultaneous segregation of many genes affecting the trait. These effects convert the intrinsically discrete variation to a continuous one. Quantitative genetics is used to distinguish between the two factors that cause continuous variability to occur.

Breeding quantitative traits

Breeding quantitative traits is more challenging than breeding qualitative traits. A discussion of quantitative genetics will give the reader an appreciation for the nature of quantitative traits and a better understanding of their breeding.

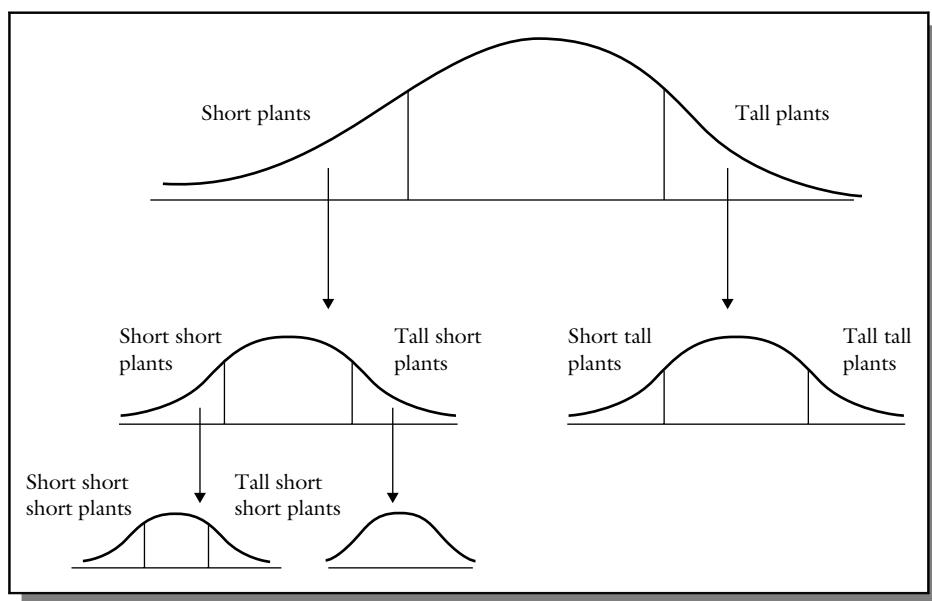


Figure 9.7 Quantitative traits are influenced to a larger degree by the environment than are qualitative traits.



11

Plant genetic resources

Purpose and expected outcomes

Human societies are constantly changing. Primitive societies change to become modern. Modern societies are more technologically advanced than their preceding counterparts. As societies evolve, their needs and preferences also change. The need for food, quantity and quality, as well as other plant products are constantly changing. The way food is produced has also changed over time. Some of these changes have been necessitated by the changes in the cultural environment as a result of humans interfering with the natural balance of nature. New pathogens have emerged in modern agriculture. Some cultivation practices adversely impact the environment. With the current environmental consciousness in society, there is a demand to reduce pesticide use. Consumer needs and preferences have changed over the years. All these changes require that plant breeders stay vigilant in developing new cultivars on a regular basis to respond to these changes. Plant breeders depend on variability (germplasm) to develop new cultivars. Some of it is obtained from local sources, while some comes from distant regions, each source with its unique advantages and disadvantages. It is important to note that, being a natural resource, germplasm is susceptible to erosion. To facilitate its use, germplasm is collected, characterized and properly stored and managed for use by plant breeders.

After completing this chapter, the student should be able to:

- 1 Discuss the importance of germplasm to plant breeding.
- 2 Describe the various sources of germplasm available to breeders, their advantages and disadvantages.
- 3 Define the term genetic vulnerability and discuss its implications to plant breeding.
- 4 Discuss the mechanisms for conservation of germplasm.
- 5 Discuss the international role in germplasm conservation.

11.1 Importance of germplasm to plant breeding

Germplasm is the lifeblood of plant breeding without which breeding is impossible to conduct. It is the

genetic material that can be used to perpetuate a species or population. It not only has reproductive value, but through genetic manipulation (plant breeding), germplasm can be improved for better performance of the crop. Germplasm provides the materials (parents)

used to initiate a breeding program. Sometimes, all plant breeders do is to evaluate plant germplasm and make a selection from existing biological variation. Promising genotypes that are adapted to the production region are then released to producers. Other times, as discussed in Chapter 5, breeders generate new variability by using a variety of methods such as crossing parents, mutagenesis (inducing mutations) and, more recently, gene transfer. This base population is then subjected to appropriate selection methods, leading to the identification and further evaluation of promising genotypes for release as cultivars. When breeders need to improve plants, they have to find a source of germplasm that would supply the genes needed to undertake the breeding project. To facilitate the use of germplasm, certain entities (germplasm banks) are charged with the responsibility of assembling, cataloguing, storing, and managing large numbers of germplasm. This strategy allows scientists ready and quick access to germplasm when they need it.

11.2 Centers of diversity in plant breeding

The subject of centers of diversity was first discussed in Chapter 10. Whereas the existence of centers of crop origin or domestication is not incontrovertible, the existence of natural reservoirs of plant genetic variability has been observed to occur in certain regions of the world. These centers are important to plant breeders because they represent pools of diversity, especially wild relatives of modern cultivars.

Plant breeding may be a victim of its own success. The consequence of selection by plant breeders in their programs is the steady erosion or reduction in genetic variability, especially in the highly improved crops. Modern plant breeding tends to focus on a small amount of variability for crop improvement. Researchers periodically conduct plant explorations (or collections) to those centers of diversity where wild plants grow in their natural habitats, to collect materials that frequently yield genes for addressing a wide variety of plant breeding problems, including disease resistance, drought resistance, and chemical composition augmentation.

11.3 Sources of germplasm for plant breeding

Germplasm may be classified into five major types – **advanced (elite) germplasm, improved germplasm,**

landraces, wild or weedy relatives, and genetic stocks. The major sources of variability for plant breeders may also be categorized into three broad groups – **domesticated plants, undomesticated plants, and other species or genera.**

11.3.1 Domesticated plants

Domesticated plants are those plant materials that have been subjected to some form of human selection and are grown for food or other uses. There are various types of such material:

- **Commercial cultivars.** There are two forms of this material – **current cultivars** and **retired or obsolete cultivars.** These are products of formal plant breeding for specific objectives. It is expected that such genotypes would have superior gene combinations, be adapted to a growing area, and have a generally good performance. The obsolete cultivars were taken out of commercial production because they may have suffered a set-back (e.g., susceptible to disease) or higher performing cultivars were developed to replace them. If desirable parents are found in commercial cultivars, the breeder has a head start on breeding since most of the gene combinations would already be desirable and adapted to the production environment.
- **Breeding materials.** Ongoing or more established breeding programs maintain variability from previous projects. These intermediate breeding products are usually genetically narrow-based because they originate from a small number of genotypes or populations. For example, a breeder may release one genotype as a commercial cultivar after yield tests. Many of the genotypes that made it to the final stage or have unique traits will be retained as breeding materials to be considered in future projects. Similarly, genotypes with unique combinations may be retained.
- **Landraces.** Landraces are farmer-developed and maintained cultivars. They are developed over very long periods and have co-adapted gene complexes. They are adapted to the growing region and are often highly heterogeneous. Landraces are robust, having developed resistance to the environmental stresses in their areas of adaptation. They are adapted to unfavorable conditions and produce low but relatively stable performance. Landraces, hence, characterize subsistence agriculture. They may be used as starting material in mass selection or pure line breeding projects.

- **Plant introductions.** The plant breeder may import new, unadapted genotypes from outside the production region, usually from another country (called plant introductions). These new materials may be evaluated and adapted to new production regions as new cultivars, or used as parents for crossing in breeding projects.
- **Genetic stock.** This consists of products of specialized genetic manipulations by researchers (e.g., by using mutagenesis to generate various chromosomal and genomic mutants).

11.3.2 Undomesticated plants

When desired genes are not found in domesticated cultivars, plant breeders may seek them from wild populations. When wild plants are used in crosses, they may introduce wild traits that have an advantage for survival in the wild (e.g., hard seed coat, shattering, indeterminacy) but are undesirable in modern cultivation. These undesirable traits have been selected against through the process of domestication. Wild germplasms have been used as donors of several important disease and insect resistance genes and genes for adaptation to stressful environments. The cultivated tomato has benefited from such introgression by crossing with a variety of wild *Licopersicon* species. Other species such as potato, sunflower, and rice have benefited from wide crosses. In horticulture, various wild relatives of cultivated plants may be used as rootstock in grafting (e.g., citrus, grape) to allow cultivation of the plant in various adverse soil and climatic conditions.

11.3.3 Other species and genera

Gene transfer via crossing requires that the parents be cross-compatible or cross-fertile. As previously stated, crossing involving parents from within a species is usually successful and unproblematic. However, as the parents become more genetically divergent, crossing (wide crosses) is less successful, often requiring special techniques (e.g., embryo rescue) for intervening in the process in order to obtain a viable plant. Sometimes, related species may be crossed with little difficulty.

11.4 Concept of gene pools of cultivated crops

J.R. Harlan and J.M.J. de Wet proposed a categorization of gene pools of cultivated crops according

to the feasibility of gene transfer or gene flow from those species to the crop species. Three categories were defined, primary, secondary, and tertiary gene pools:

- (i) **Primary gene pool (GP1).** GP1 consists of biological species that can be intercrossed easily (interfertile) without any problems with fertility of the progeny. That is, there is no restriction to gene exchange between members of the group. This group may contain both cultivated and wild progenitors of the species.
- (ii) **Secondary gene pool (GP2).** Members of this gene pool include both cultivated and wild relatives of the crop species. They are more distantly related and have crossability problems. Nonetheless, crossing produces hybrids and derivatives that are sufficiently fertile to allow gene flow. GP2 species can cross with those in GP1, with some fertility of the F1, but more difficulty with success.
- (iii) **Tertiary gene pool (GP3).** GP3 involves the outer limits of potential genetic resources. Gene transfer by hybridization between GP1 and GP3 is very problematic, resulting in lethality, sterility, and other abnormalities. To exploit germplasm from distant relatives, tools such as embryo rescue and bridge crossing may be used to nurture an embryo from a wide cross to a full plant and to obtain fertile plants.

A classification of dry bean and rice is presented in Figure 11.1 for an illustration of this concept. In assembling germplasm for a plant breeding project, the general rule is to start by searching the domesticated germplasm collection first, before considering other sources, for reasons previously stated. However, there are times when the gene of interest occurs in undomesticated germplasm, or even outside the species. Gene transfer techniques enable breeders to transfer genes beyond the tertiary gene pool. Whereas all crop plants have a primary gene pool that includes the cultivated forms, all crops do not have wild forms in their GP1 (e.g., broad bean, cassava, and onions whose wild types are yet to be identified). Also, occasionally, the GP1 may contain taxa of other crop plants (e.g., almond belongs to the primary gene pool of peach). Most crop plants have a GP2, which consists primarily of species of the same genus. Some crop plants have no secondary gene pools (e.g., barley, soybean, onion, broad bean).

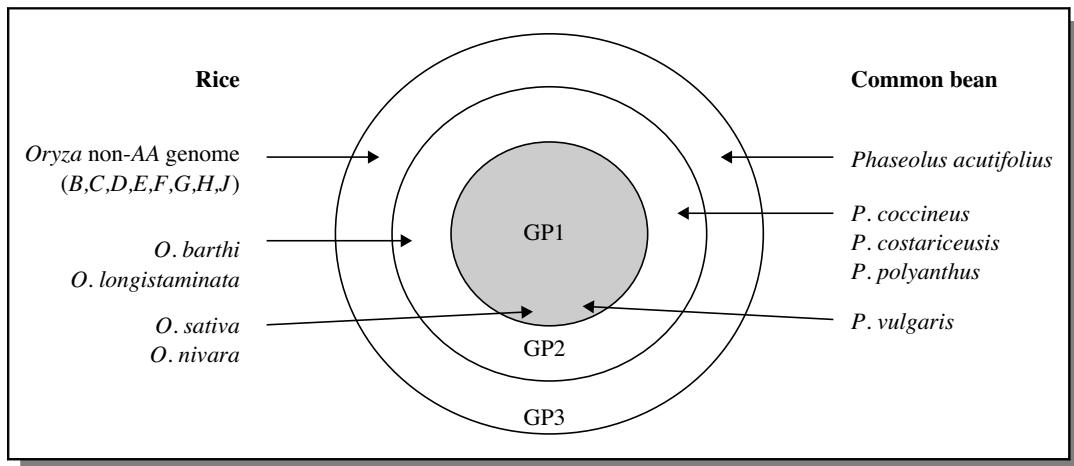


Figure 11.1 Crop gene pools. Harlan proposed the crop gene pools to guide the germplasm use by plant breeders. The number of species in each of the pools that plant breeders are using varies among crops. Harlan suggested that breeders first utilize the germplasm in the GP1 and proceed outwards.

11.5 Concept of genetic vulnerability

Genetic vulnerability is an important issue in modern plant breeding, brought about largely by the manner in which breeders go about developing new and improved cultivars for modern society.

11.5.1 What is genetic vulnerability?

Genetic vulnerability is a term used to indicate the genetic homogeneity and uniformity of a group of plants that predisposes it to susceptibility to a pest, pathogen, or environmental hazard of large-scale proportions. It is a complex problem that involves issues such as crop evolution, trends in breeding, trends in biological technology, decisions by crop producers, demands and preferences of consumers, and other factors. As a result of a combination of the above factors, a certain kind of crop cultivar (genotype) is developed for the agricultural production system. A case in point is the 1970 epidemic of southern leaf blight (*Helminthosporium maydis*) in the United States that devastated the corn industry. This genetic vulnerability in corn was attributed to uniformity in the genetic background in corn stemming from the widespread use of the T-cytoplasm in corn hybrid seed production.

Prior to the use of the T-cytoplasm, hybrid corn breeders had to detassel their non-pollen parent. The T-cytoplasm contained a gene in the mitochondrion

(mtDNA) that conferred male sterility upon the plant, eliminating the need for detasseling. However, a male fertility restorer gene in parent B was needed to restore fertility to the F₁ (AxB) seed. Large acreages of hybrids developed this way were grown. Unfortunately, a new strain of *H. maydis* surfaced around 1970 with devastating consequences due to the large acreage of the crop.

Another classic case of crop vulnerability occurred in Europe. The late blight of potato was introduced from the Americas into Europe around 1845. It quickly spread in the area, causing major epidemics, in Ireland especially. The consequence of this disaster included starvation, disease, death, and the emigration of over five million Irish nationals to the United States.

Genetic uniformity and the consequence of vulnerability to a large extent are created, maybe inadvertently, in response to the preference of both consumers and producers for uniform products in some situations. Further, it appears the key factor is the gene shared in common rather than the genetic background in which the gene occurs. For example, a gene that may trigger a disease epidemic in one species may occur in another species. Consequently, when there is a disease outbreak, it is not confined but spreads to different species. *Phenotypically* dissimilar crops can share a trait that is simply inherited and that predisposes them to susceptibility to an adverse biotic or abiotic factor. A case in point is the

chestnut blight (*Cryphonectria parasitica*) epidemic that occurred in the United States in which different species of the plant were affected.

It should be pointed out that genetic vulnerability is an issue in hindsight, since the disaster it causes cannot be predicted. For example, the use of the T-cytoplasm was a great strategy until the southern leaf blight appeared in the corn fields. Similarly, the *mlo* mildew resistance in barley against powdery mildew is beneficial to farmers in Europe, where about 70% of all spring barley incorporated that gene. However, should an unforeseen event similar to the southern leaf blight occur, would this once outstanding achievement be viewed in the negative light? The message is that breeders should consider diversity when selecting parents for breeding programs so that crop vulnerability would be reduced.

Germplasm in which insufficient diversity is significant includes grapes, sweet potato, cucurbits, tropical fruits and nuts, cool-season food legumes, peach, cherry, walnuts, herbaceous and woody ornamentals.

11.5.2 Key factors in genetic vulnerability induced crop failure

The key factors that are responsible for the disastrous epidemics attributable to genetic vulnerability of crops are:

- The desire by growers and consumers for uniform and blemish-free products, which is often achieved by breeders by applying time and again the same (sets of) gene(s).
- The acreage devoted to the crop cultivars carrying such popular genes and method of field production.

Where the cultivars with the susceptible gene are widely distributed in production (i.e., most farmers use the same cultivars), the risk of disaster will be equally high but unpredictable! Further, where the threat is biotic, the mode of dispersal of the causal agent and the presence of a favorable environment for the pathogen to develop would increase the risk of disasters (e.g., wind mode of dispersal of spores or propagules will cause a rapid spread of the disease). In biotic disasters, the use of a single source of resistance to the pathogen is perhaps the single most important factor in vulnerability. However, the effect can be exacerbated by practices such as intensive and continuous monoculture using one cultivar or various cultivars

carrying the same cytoplasm or resistance genes. Under such production practices, the pathogen only has to overcome one resistance gene, resulting in rapid disease advance and great damage to crop production.

11.6 What plant breeders can do to address crop vulnerability

The issue of the importance of genetic vulnerability is best left to plants to decide.

11.6.1 Reality check

First and foremost, plant breeders need to be convinced that genetic vulnerability is a real and present danger. Without this first step, efforts to address the issue are not likely to be taken seriously. A study by D.N. Duvick in 1984, albeit dated, posed the question “How serious is the problem of genetic vulnerability in your crop?” to plant breeders. The responses by breeders of selected crops (cotton, soybean, wheat, sorghum, maize, and others) indicated a wide range of perception of crop vulnerability, ranging from 0 to 25% thinking it was serious, and 25–60% thinking it was not a serious problem (at least at that time). Soybean and wheat breeders expressed the most concern about genetic vulnerability. Their fears are most certainly founded since, in soybean, it is estimated that only six cultivars constitute more than 50% of the crop acreage of North America. Similarly, more than 50% of the acreage of many crops in the United States is planted to less than 10 cultivars per crop.

An early assessment of genetic uniformity of major crops in the United States by the National Research Council showed that, in peanuts, 95% of the acreage was dominated by nine major cultivars. In corn, 71% of the acreage was sown to only six major varieties, while in the case of dry bean 60% of the acreage was cultivated to only two major cultivars. The message here is that several major crops in the United States are vulnerable to epidemics. However, there is reversal in the longtime trend toward increased diversification.

11.6.2 Use of wild germplasm

Many of the world’s major crops are grown extensively outside the centers of origin where they

co-evolved with pests and pathogens. Breeders should make deliberate efforts to expand the genetic base of their crops by exploiting genes from the wild progenitors of their species that are available in various germplasm repositories all over the world.

11.6.3 Paradigm shift

As D. Tanksley and S.R. McCouch of Cornell University point out, there is a need for a paradigm shift regarding the use of germplasm resources. Traditionally, breeders screen accessions from exotic germplasm banks on phenotypic basis for clearly defined and recognizable features of interest. Desirable genotypes are crossed with elite cultivars to introgress genes of interest. However, this approach is effective only for the utilization of simply inherited traits (conditioned by single, mostly dominant genes). The researchers proposed a shift from the old paradigm of looking for phenotypes to a new paradigm of looking for genes. To accomplish this, the modern techniques of genomics may be used to screen exotic germplasm by a gene-based approach. They proposed the use of molecular linkage maps and a new breeding technique called **advanced backcross QTL** introgression that allows the breeder to examine a subset of genes from the wild exotic plant in the genetic background of an elite cultivar.

11.6.4 Use of biotechnology to create new variability

The tools of modern biotechnology, such as recombinant DNA, cell fusion, somaclonal variation, and others, may be used to create new variability for use in plant breeding. Genetic engineering technologies may be used to transfer desirable genes across natural biological barriers.

11.6.5 Gene pyramiding

Plant breeders may broaden the diversity of resistance genes as well as introducing multiple genes from different sources into cultivars using the strategy of **gene pyramiding**, that is, introducing more than one resistance gene into one genotype. This approach will reduce the uniformity factor in crop vulnerability.

11.7 Wild (exotic) germplasm in plant breeding

Domestication has narrowed the genetic basis of many modern crop species. Breeders are returning to the reservoirs of genetic variations that were left behind by the process of domestication and thus exist in the wild, as well as in primitive germplasm such as landraces. Using wild germplasm in plant breeding has many challenges. It is time consuming and laborious; the unadapted germplasm has many undesirable genes; cross incompatibility between the wild and cultivated species often occurs; a successful cross often results in F₁ hybrid sterility and subsequent infertility of the segregating generations. Further, crosses exhibit reduced recombination between the wild and cultivated chromosomes as well as linkage drag (tight linkage of trait of interest in the wild genome to undesirable genes which are transferred to the cultivated genome).

Notwithstanding these and other challenges, plant breeders have used the method of introgression breeding to transfer agriculturally valuable traits from the wild into commercial cultivars. However, most of the successes have been achieved in the transfer of monogenic traits (e.g., disease resistance) leaving many other major traits like yield, quality, response of abiotic stress, and others relatively unattended to because of the complex genetics that underlie their inheritance (quantitative traits). These complex traits are controlled by quantitative trait loci (QTLs) which are being characterized through the use of modern molecular technologies.

The success and effectiveness of introgression of disease resistance genes into crop species from wild relatives varies by crop. Factors to consider include the amount of diversity within the crop species, ease of hybridization with wild relatives, and the complexity of the genetic control of the trait. Some crop breeders (e.g., tomato breeders) use genes from wild relatives more frequently than other breeders, such as sorghum breeders, who appear to find their needs in the domesticated crop species. The recurrent back-crossing approach previously mentioned has been used to improve over a dozen commercial cultivars of tomato. In tomato, wild relatives provided genes for enhancing the nutritional value (vitamin C and beta-carotene) and solids content, significantly boosting the commercial value of the crop. A specific example is the transfer of a gene, *B*, from the wild tomato,

Lycopersicon pennellii, which increased the level of provitamin A (beta-carotene) in the fruit by over 15-fold.

The impact of introgression of genes from the wild into adapted cultivars has been dramatic in some crops. For example, the resistance to the devastating late blight of potato was found in a wild species. Similarly, resistance to the root knot nematode in peanut was obtained from three wild species. A wild relative of rice, *Oryza nivara*, growing in the wild in Uttar Pradesh, India, was found to have one single gene for resistance to the grassy stunt virus, a disease that devastated the crop in South and South East Asia in the 1970s. In wheat, yield improvements associated with the transfer of a chromosomal segment carrying a rust resistance gene (*Lr19*) from *Agropyron elongatum* (tall wheat grass) has been reported. Similarly, a high-grain protein QTL from *Triticum dicoccoides*

(wild emmer wheat) was introgressed into cultivated wheat to improve pasta flour quality.

In spite of these and other successes with introgression of wild genes into cultivated species, the use of this approach is anything but routine in plant breeding. Dani Zamir proposed the development of **exotic libraries** to facilitate and accelerate the exploitation of wild germplasm in breeding (Figure 11.2). His approach centers on developing a library of introgressed lines to be made available to breeders. At the present, should a breeder desire to re-screen the progeny of an interspecific cross that had been previously generated for a new trait, one would have to start from the parents to develop the required generation.

Even though modern technology has made it possible to develop a large number of informative genetic markers and high-density marker maps for

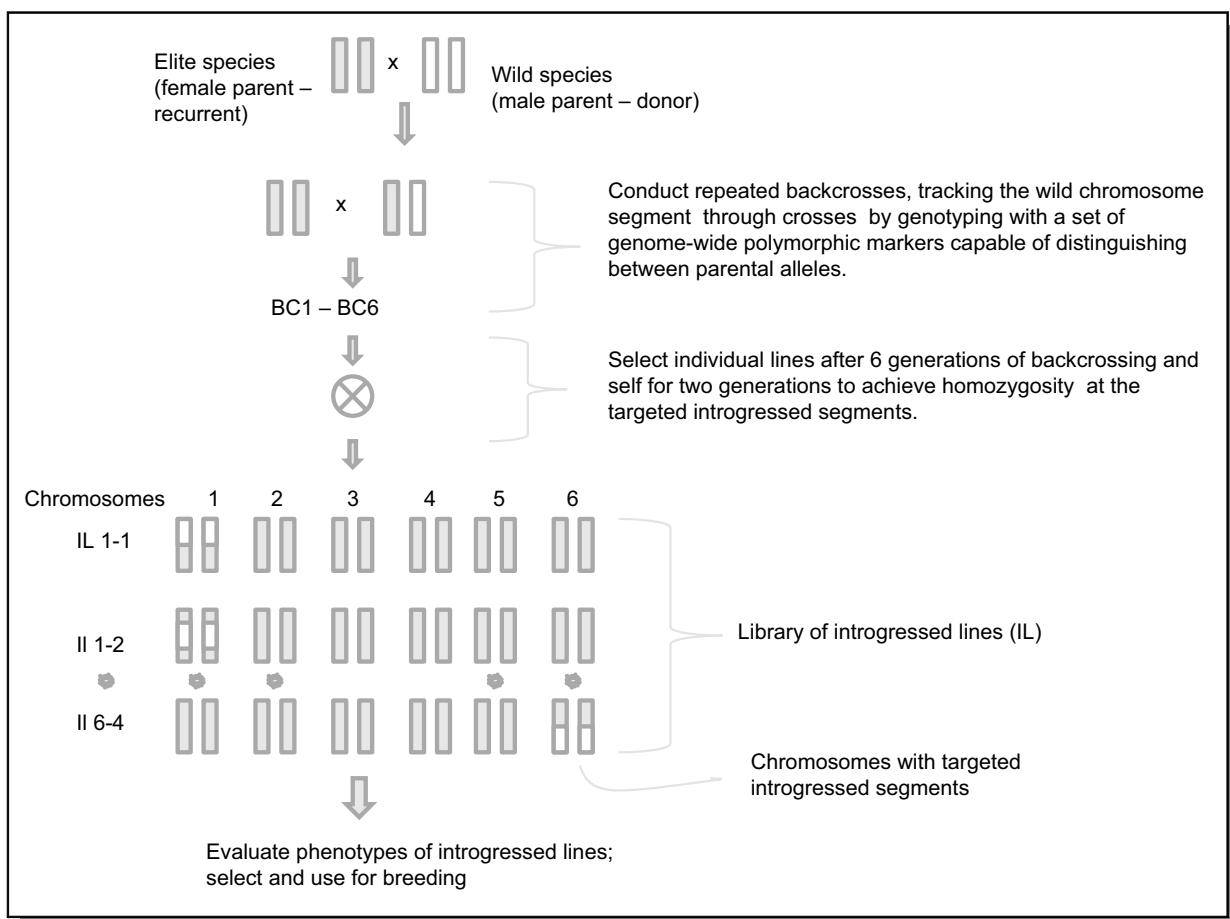


Figure 11.2 Generating and screening an exotic genetic library.

monogenic and QTLs that affect phenotypic variation, the problem still remains that in F_2 , backcross, advanced backcross, and recombinant inbred populations resulting from divergent genomes, genetic analysis remains challenging because the presence of large proportion of wild genomes renders these populations partially sterile. Consequently, their use in identification of QTLs for improving yield and other important agronomic traits is problematic. Creating an exotic library would solve this problem.

An exotic library essentially consists of a set of lines, each carrying a single defined chromosome segment from a donor species in an otherwise uniform elite genetic background. It would take about 10 generations to create these lines, which would become immediately available for direct use in breeding. The lines would be excellent tools for detecting and mapping agronomic traits. Because the lines differ from the elite variety by only a single defined chromosomal segment, they would resemble the cultivated variety and have reduced sterility problems.

11.8 Plant genetic resources conservation

Plant breeders manipulate variability in various ways; for example, they assemble, recombine, select, and discard. The preferential use of certain elite genetic stock in breeding programs has narrowed the overall genetic base of modern cultivars. As already noted, pedigree analysis indicates that many cultivars of certain major crops of world importance have common ancestry, making the industry vulnerable to disasters (e.g., disease epidemic, climate changes). National and international efforts have been mobilized to conserve plant genetic resources.

11.8.1 Why conserve plant genetic resources?

There are several reasons why plant genetic resources must be conserved:

- Plant germplasm is exploited for food, fiber, feed, fuel, and medicines by agriculture, industry, and forestry.
- As a natural resource, germplasm is a depletable resource.
- Without genetic diversity, plant breeding cannot be conducted.

- Genetic diversity determines the boundaries of crop productivity and survival.
- To cater for future needs.

As society evolves, its needs will keep changing. Similarly, new environmental challenges might arise (e.g., new diseases, abiotic stresses, mechanical harvesting due to rise in labor costs) for which new variability might be needed for plant improvement.

When a genotype is unable to respond fully to the cultural environment, or able to resist unfavorable conditions thereof, crop productivity diminishes. The natural pools of plant genetic resources are under attack from the activities of modern society – urbanization, indiscriminate burning, clearing of virgin land for farming, to name a few. These and other activities erode genetic diversity in wild populations. Consequently, there is an urgent need to collect and maintain samples of natural variability. The actions of plant breeders also contribute to genetic erosion as previously indicated. High-yielding narrow genetic based cultivars are penetrating crop production systems all over the world, displacing the indigenous high-variability landrace cultivars. Some 20 000 plant species are suggested to be endangered.

11.8.2 Genetic erosion

Genetic erosion may be defined as the decline in genetic variation in cultivated or natural populations largely through the action of humans. Loss of genetic variation may be caused by natural factors, and by the actions of crop producers, plant breeders, and others in the society at large.

- **Natural factors.** Genetic diversity can be lost through natural disasters such as large-scale floods, wild fires, and severe and prolonged drought. These events are beyond the control of humans.
- **Action of crop producers.** Right from the beginnings of agriculture, farmers have engaged in activities that promote genetic erosion. These include clearing of virgin land in, especially, germplasm-rich tropical forests and the choice of planting material (narrow genetic base cultivars). Farmers, especially in developed economies, primarily grow improved seed that are genetically pure, having replaced most or all landraces with these superior cultivars. Also, monoculture, instead of mixed cropping, tends promotes the narrowing of genetic diversity as large

tracts of land are planted to uniform cultivars. Extending grazing lands into wild habitats by livestock farmers destroys wild species and wild germplasm resources.

- **Action of breeders.** Farmers plant what breeders develop, and breeders develop what farmers and consumers and industry demand. Some methods used for breeding (e.g., pure lines, single cross, multilines) promote uniformity and narrower genetic base. When breeders find superior germplasm, the tendency is to use it as much as possible in cultivar development. In soybean, as previously indicated, most of the modern cultivars in the United States can be traced back to about half a dozen parents. This practice causes severe reduction in genetic diversity.
- **Problems with germplasm conservation.** In spite of good efforts by curators of germplasm repositories to collect and conserve diversity, there are several ways in which diversity in their custody may be lost. The most obvious loss of diversity is attributed to human errors in the maintenance process (e.g., improper storage of materials leading to loss of germination, and hence loses off accessions and the variation they represent). Also, when germplasm is planted in the field, natural selection pressure may eliminate some unadapted genotypes. Hybridization, as well as genetic drift cause shifts in allele frequencies in small populations. They occur as consequences of periodic multiplication of the germplasm holdings by curators.
- **General public action.** As previously indicated, there is an increasing demand on land with increasing population. Such demands include settlement of new lands and the demand for alternative use of the land (e.g., for recreation, industry, roads) to meet the general needs of modern society. These actions tend to place wild germplasm in jeopardy. Such undertakings often entail clearing of virgin land where wild species occur.

11.8.3 Selected impact of germplasm acquisition

Impact on North American agriculture

Very few crops have their origin in North America. It goes without saying that North American agriculture owes its tremendous success to plant introductions, which brought major crops such as wheat, barley, soybean, rice, sugar cane, alfalfa, corn, potatoes, tobacco, and cotton to this part of the world. North America currently is the world's leading producer of many of

these crops. Spectacular contributions by crop introductions to US agriculture include the following:

- Avocados – Introduced in 1898 from Mexico, this crop has created a viable industry in California.
- Rice – Varieties introduced from Japan in 1900 laid the foundation for the present rice industry in Louisiana and Texas.
- Spinach – A variety introduced from Manchuria in 1900 is credited with saving the Virginia spinach industry from blight disaster in 1920.
- Peach – Many US peach orchards are established by plants growing on root stalks obtained from collections in 1920.
- Oat – One of the world's most disease-resistant oat varieties was developed from germplasm imported from Israel in 1960s.

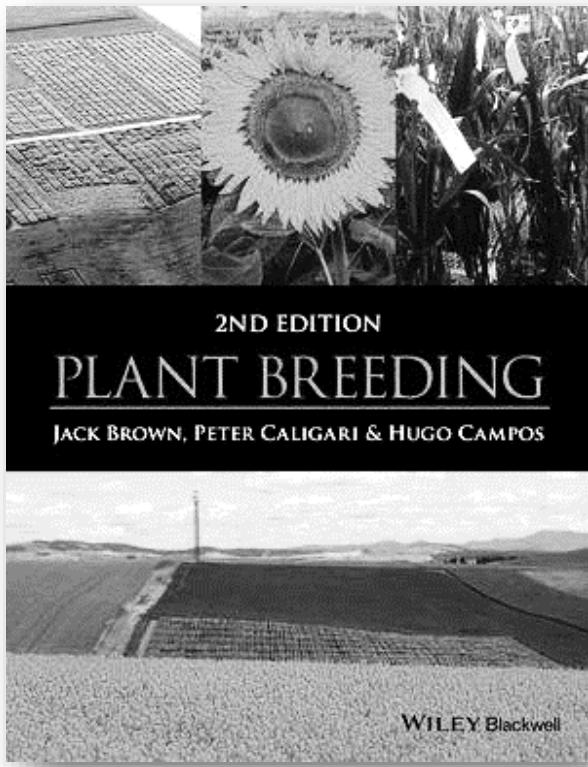
Other parts of the world

A few examples include dwarf wheats introduced into India, Pakistan, and Philippines as part of the Green Revolution, and introduction of soybeans and sunflower into India, have benefited the agriculture of these countries.

11.9 Nature of cultivated plant genetic resources

Currently, five kinds of cultivated plant materials are conserved by concerted worldwide efforts – **landraces (or folk, primitive varieties), obsolete varieties, commercial varieties (cultivars), plant breeder's lines, and genetic stocks**. Landraces are developed by indigenous farmers in various traditional agricultural systems or are products of nature. They are usually very variable in composition. Obsolete cultivars may be described as “ex-service” cultivars because they are no longer used for cultivation. Commercial cultivars are elite germplasm currently in use for crop production. These cultivars remain in production usually from 5–10 years before becoming obsolete and replaced. Breeder's lines may include parents that are inbred for hybrid breeding, genotypes from advanced yield tests that were not released as commercial cultivars, and unique mutants. Genetic stocks are genetically characterized lines of various species. These are advanced genetic materials developed by breeders, and are very useful and readily accessible to other breeders.

2.2 Study book



Title:	<i>Plant Breeding</i>
Edition:	2
Authors:	<i>Jack Brown, Peter Caligari, and Hugo Campos</i>
Year:	2014
Publisher:	<i>John Wiley & Sons Inc</i>

2.2.1.-2.2.2

Brown et al. (2014)
Page 40-65, and 125-147

4

Breeding Schemes

4.1 Introduction

All successful breeding programmes have been designed around a breeding scheme. The breeding scheme determines the passage of breeding lines through the selection process, and through to the increase of planting material for cultivar release. The process of selection will be carried out over a number of years, and under differing environmental conditions. The early selection stages of breeding programmes will involve screening many thousands of different genotypes. The early screening is therefore relatively crude, and in many instances this involves only visual selection. After each round of selection, the '*better*', more adapted, or more disease-resistant genotypes will be retained for further evaluation while the least adapted lines will be discarded. This process will be repeated over a number of years, and at each stage the number of individual genotypes or populations is reduced and evaluation is conducted with greater precision in estimating the worth of each entry.

The breeding scheme used will be highly dependent upon the crop species and the type of cultivar (pure-line/self-pollinated, open-pollinated, hybrid, clonally propagated, synthetic, etc.) that is being developed. The general philosophy for developing a clonal cultivar like potato is therefore different from a pure-line cereal cultivar, say barley. In the former, breeding selections are genetically fixed through vegetative propagation, but there will be a low rate of multiplication of planting material. In

the latter, there will be more rapid increase of planting material, although the segregating nature of the early-generation breeding lines will complicate the selection process.

The most effective breeding schemes will utilize the positive attributes of a crop species while minimizing difficulties that might arise through the selection process. In the following section the general breeding schemes for pure-line, open-pollinated, hybrid and clonally propagated cultivars will be explained, along with mention of the schemes used for developing multilines and synthetics.

4.2 Development of pure-line cultivars

Crops that are generally produced as pure-line cultivars include barley, chickpea, flax, lentil, millet, peas, soybean, tobacco, tomato and wheat.

One and a half centuries ago, most inbred crop species were grown in agriculture as 'landraces'. Landraces were locally grown populations which were, in fact, a collection of many different genotypes grown in mixture and which were, of course, both genetically and phenotypically variable. Pure-line (synonymous with inbred) cultivars were developed first from these landraces by farmers who selected specific (presumably more productive or least infected with disease or pests) plants from the mixed populations and maintained these in isolation, thus encouraging selfed

progenies, and eventually developed homozygous, or near-homozygous, lines. It is reasonable to assume that these homozygous lines were indeed more productive than the original landraces because by the end of the 19th century, landraces had almost completely disappeared in countries with advanced agricultural systems.

These early ‘pure-line breeders’ used the naturally existing genetic variation within the landraces they were propagating and the natural tendency of some species to self-pollinate (e.g. wheat) at a high frequency. However, this strategy has a limited potential in terms of generating new variation, and so modern plant breeders have to continuously generate genetic variation and hence the three-phase breeding schemes were established to **create genetic variation, identify desirable recombinant lines within progenies and stabilize and increase the desired genotype**. It is interesting to note, however, that recently a number of plant breeders have returned to old landraces of wheat and barley to examine their wealth of genetic diversity, as well as to testing combinations of lines in ‘modern’ landrace combinations (i.e. multilines). Unfortunately most of the landraces that existed even 100 years ago are no longer available and potentially valuable germplasm and adapted combinations have been lost.

By far the most commonly used method of generating genetic variation within inbreeding species is via sexual reproduction using artificial hybridization. There are, of course, other ways to produce genetic variation. For example, variation can be produced by induced mutation, somatic variation, somatic hybridization and recombinant DNA techniques (all discussed in later chapters).

After sexual crossing between genotypes and then selfing them to generate suitable genetic variation, plant breeders will then traditionally screen the segregating population for desirable ‘segregants’ while continuing to self-pollinate successive generations, to produce homozygous lines. Thus, accomplishing the last two steps of the breeding scheme (selection and stabilization) can be achieved more or less simultaneously.

4.2.1 Homozygosity

One of the difficulties in selecting desirable recombinant lines in pure-line breeding is related to

segregating populations and the masking of specific character expression as a result of the dominant/recessive nature of the segregating alleles in the heterozygotes. Another consideration is the relationship between genetic homozygosity and ‘commercial inbred lines’. The definition of complete homozygosity is that all the alleles at all loci are identical by descent, that is, there is no heterozygosity at **any** locus. However, for practical commercial exploitation, the level of homozygosity does not need to be complete. Clearly the lines must basically breed ‘true to type’ but this is by no means absolute. The degree of homozygosity is determined by the level of inbreeding and directly related, for example, to the number of selfing generations that have been performed. Consider the simple case of just one locus with two possible alleles A - a :

Parents	AA × aa			Frequency of heterozygotes		
F ₁	Aa			100%		
F ₂ Frequency	AA 1/4	Aa 1/2	aa 1/4	50%		
F ₃ Frequency	AA 1/4	AA 1/8	Aa 1/4	aa 1/8	aa 1/4	25%

Consider now a more complex situation where six loci are involved, as set out below. Of these six loci, two, loci A and F , have the same allele in both parents (and are both homozygous) and so the F₁ is also homozygous at these two loci. At the other four loci the parents are homozygous but have different alleles, and so the F₁ is heterozygous at these loci and subsequently these segregate in the F₂.

Parents	$AAbbCCDDeeff \times AABbCcddEEff$
F ₁	$AABbCcDdEeff$
F ₂	$AABBCCDdeeff, AABbCcDDEeff,$ $AAAbCCEeDdff, AABbCcDdEeff$, etc.

This can be generalized in mathematical terms as follows. Consider an F₁ that is heterozygous at n loci; heterozygosity (h) at any single loci after g generations ($g = 0$ at F₁) of selfing will be:

$$h = (1/2)^g$$

Therefore the probability (p) of homozygosity at all n loci will be:

$$p = (1 - h)^n$$

Hence after g generations:

$$p = [1 - (1/2)^g]^n$$

This can also be written as:

$$p = [(2^g - 1)/2^g]^n$$

The level of homogeneity (i.e. uniformity in appearance/phenotype as opposed to genetic homozygosity) required in an inbred cultivar will depend to a varying extent on the personal choice of the breeder, seed regulatory agencies, farmers and end-users. Almost all pure-line breeding schemes involve selection of individual plants at one or more stages in the breeding scheme. The stages of single plant selection will have a large impact regarding the degree of heterogeneity in the end cultivar. If single plant selection is carried out at an early generation, say F_2 , there may be greater heterogeneity within the resulting cultivar compared with a situation where single plant selection was delayed until a later generation, say F_8 , when individual plants would be more homozygous. Breeders must ensure that a level of uniformity and stability exists throughout multiplication and into commercialization. Farmers and processors (such as millers) will have preferences for cultivars that are homozygous, and hence homogeneous, for particular characters. These characters may be related to uniform maturity, plant height or other traits related to ease of harvest. Many believe that farmers are not concerned with uniformity of characters that do not interfere with end-use performance (e.g. flower colour segregation). However, farmers take a natural pride in their farms and, therefore, like to grow '*nice-looking*' crops, and these are ones which are uniform for almost all visible characters. In any case they certainly like uniformity in terms of characters such as harvest date! For most end-users there will be an obvious preference for cultivars with high uniformity of desirable quality characters. For example, there may be a premium for uniform germination in malting barley or more uniform characters relating to breadmaking in wheat.

In contrast, some breeders of pure-line cultivars like to maintain a relatively high degree of heterogeneity in their developed cultivars. They believe

that this heterogeneity can help to '*buffer*' the cultivar against changes in environment and hence make the cultivar more *stable* over different environments. Often statutory authorities determine the degree of variability that is allowed in a cultivar. For example, all pure-line cultivars released in the European Union countries, along with Canada and Australia, must comply to set standards for ***distinctness, uniformity*** and ***stability*** (DUS) in Statutory National Variety Trials. In these cases it is common to have almost total homogeneity and homozygosity in released inbred cultivars.

For most breeders, time is at a premium. Therefore, some methods are commonly used to reduce the time taken to achieve homozygosity, and these include ***single seed descent*** and the use of ***off-season sites*** (this excludes the production of homozygous lines through doubled haploids, which is relevant here but will be discussed separately in a later chapter).

Single seed descent

Single seed descent involves repeatedly growing a number of individuals initiating from a segregating population, usually under high-density, low-fertility conditions, to accelerate seed-to-seed time. At maturity, a single seed from the natural self of every plant is taken and replanted. This operation is repeated a number of times to obtain homozygous plants. Single seed descent is most suited for rapid generation increase in a greenhouse where a number of growth cycles may be possible each year. Single seed descent in canola, wheat and barley can be further accelerated by growing plants under stress conditions of high density, high light, restricted root growth, and low nutrient levels, which result in stunted plants with only one or two seeds each, but in a shortened growing period compared with growth under normal conditions (up to three or four generations in a year are possible in barley or spring canola).

It is very important, however, when using single seed descent, that unintentional selection is not being carried out for adverse characters. For example, in a single seed descent scheme in winter wheat (where plants will require a vernalization period prior to initiating a reproductive phase), vernalization requirements may be overcome artificially in a cold room. If this is done, then care should

be taken so that all seedlings do indeed receive sufficient cold treatment to overcome the vernalization requirement – otherwise the system will select the plant types with lower vernalization requirements. In addition, some genetic characteristics are not fully expressed when plants are grown under high competition stress conditions used for single seed descent. For example, the *erectoides* dwarfing gene (*ert*) in barley plants is not expressed under single seed descent in the glasshouse, and therefore genotypes cannot be selected for dwarfism under these conditions. In any case it is strongly advised that **no** selection be practised during this phase.

Off-season sites

Off-season growing sites can also reduce the time for achieving a desired level of homozygosity. This is possible by having more than one growing season per year. Dual locations at similar latitudes in the northern and southern hemispheres are frequently used to increase either seed quantity or reduce heterozygosity in many breeding schemes. The use of off-season sites is often restricted to annual spring crops, and there are only a few good examples where they have helped accelerate homozygosity in winter annuals, and virtually none in breeding biennials.

If off-season sites are to be incorporated into the breeding scheme, care must be taken to ensure that '*selectional adaptability*' of the off-season site does not have adverse effects on the segregating plant populations. For example, the spring barley breeding scheme at the Scottish Plant Breeding Station used to increase F_4 breeding selections to F_5 by growing these lines over winter in New Zealand. Although New Zealand has a climate that is very similar to that found in Scotland, there is a completely different spectrum of races of powdery mildew. As a result, mildew-resistant selections made in New Zealand were of no relevance when grown in Scotland, and so meant that all New Zealand trials needed careful spraying to avoid powdery mildew being confounded with other performance characters. Breeding companies developing varieties for the US market have developed large networks of off-season sites around the globe for crops such as maize and soybean.

The use of off-season sites benefits farmers as they significantly reduce the time required to take new

varieties to the market, and so to respond to the ever-changing needs of humankind.

4.2.2 Breeding schemes for pure-line cultivars

There are probably as many different breeding schemes used by breeders of pure-line cultivars as there are breeders of inbreeding species. There are, however, three basic schemes: ***bulk methods***, ***pedigree methods*** and ***bulk/pedigree methods***. It should be noted that all the breeding schemes described involve more than a single cross at the *crossing stage*. A number of these crosses will be two-parent crosses (female parent \times male parent, say $P_1 \times P_2$, although many breeders use three- and four-way parent cross combinations ($[P_1 \times P_2] \times P_3$, and $[P_1 \times P_2] \times [P_3 \times P_4]$, respectively).

Bulk method

The outline of a bulk scheme is illustrated in Figure 4.1. In this scheme, genetic variation is created by artificial hybridization between chosen parents.

The F_1 and several subsequent generations, in the illustration up to and including the F_5 generations,

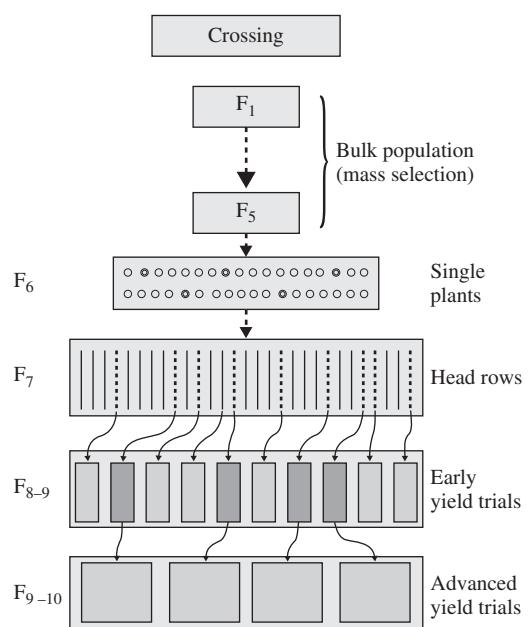


Figure 4.1 Outline of a bulk breeding scheme used for breeding pure-line crop species.

are grown as bulk populations. These bulks are left to set seed naturally which, with these species, will mean predominantly self-pollination. No conscious selection is imposed in these generation, and it is assumed that the genotypes most suited to the environment in which the bulk populations are grown will leave more offspring and hence predominate in future generations. Similarly, these bulk populations are usually grown under the stress and disease pressures common to the cultivated crop, and it is assumed that the frequency of adapted genotypes in the population increases. It is therefore very important that the bulks are grown in a suitable and representative environment. After a number of rounds of bulk increase, individual plants showing desirable characteristics are selected, often at the F₆ stage. From each selected plant, a head of seed is taken and grown as a row (a head-row). The produce from the best head-rows are bulk harvested, for initial yield trials. More advanced yield trials are grown from bulk harvest of desirable individuals.

The major advantage of the bulk method is that conscious selection is not attempted until plants have been selfed for a number of generations and hence the plants are nearly homozygous. This avoids the difficulty of selection among segregating populations where phenotypic expression will be greatly affected by levels of dominance in the heterozygotes. This method is also one of the least expensive methods of producing populations of inbred lines. The disadvantage of this scheme is the relatively long time from initial crossing until yield trials are grown. In addition, it has often been found that the natural selection, which occurs through bulk population growth, is not always that which is favourable for growth in agricultural practice. In addition, natural selection can, of course, only be effective in environments where the character is expressed. This often prevents the use of bulk methods at off-season sites.

Other methods have been used to produce homozygosity in bulk breeding schemes. These include single seed descent and doubled haploidy. Breeding schemes that use these techniques have increased the popularity of bulk breeding scheme in recent years, as the time from crossing to evaluation can be minimized. However, the basic philosophy is similar, being to produce near-homozygous lines, and thereafter select amongst these. Where rapid acceleration to homozygosity techniques are used,

it is essential to ensure that no negative selection occurs. For example, research has shown that creating homozygous breeding lines of canola (*B. napus*) through pollen culture produces a higher than random frequency of plants with low erucic acid in the seed oil. If low erucic acid content is desired, this poses a selection advantage. If, however, an industrial oilseed cultivar were desired (one with high erucic acid content), then using embryogenesis would be detrimental.

Pedigree method

The outline of a pedigree breeding scheme is shown in Figure 4.2. In a pedigree breeding scheme, single plant selection is carried out at the F₂ through to F₆ generations. The scheme begins by hybridization between chosen homozygous parental lines, and segregating F₂ families are obtained by selfing the heterozygous F₁s. Single plants are selected from amongst the segregating F₂ families. The produce from these selected plants is grown in plant/head rows at the F₃ generation. A number of the 'most desirable' single plants (in Figure 4.2, four plants)

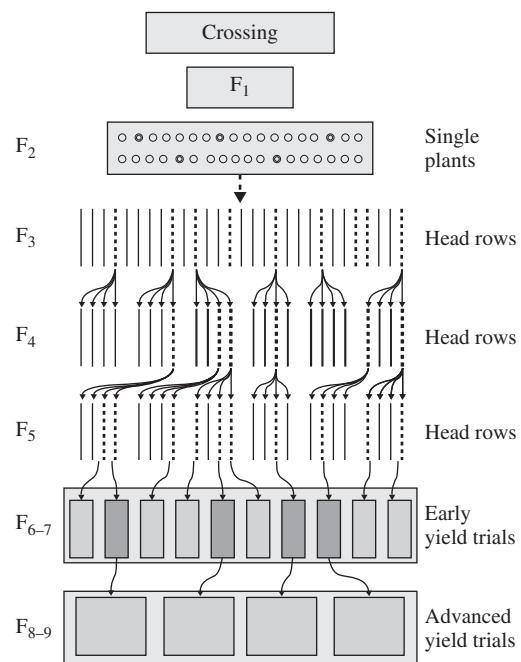


Figure 4.2 Outline of a pedigree breeding scheme used for breeding pure-line crop species.

are selected from the ‘better’ plant rows and these are grown in plant rows again at the F₄ stage. This process of single plant/head selection is repeated until plants are ‘near’ homozygous (i.e. F₆). At this stage the most productive rows are bulk harvested and used as the seed source for initial yield trials at F₇.

In addition to being laborious (as a considerable amount of record-keeping is required) and relatively expensive, annual discarding may lead to the loss of valuable genotypes, particularly under the changing environmental conditions from year to year, making selection difficult. Other disadvantages of the pedigree method are that it requires more land and labour than other methods; experienced staff with a ‘good breeders’ eye’ are necessary to make plant selections; selection is carried out on single plants where errors of observation may be great; while actual yield testing is not possible in the early generations.

If selection was effective on a single plant basis, pedigree breeding schemes would allow inferior genotypes to be discarded early in the breeding scheme, without the need for being tested in more extensive, and costly, yield trials. Unfortunately, pedigree breeding schemes offer little opportunity to select for quantitatively inherited characters, and even single gene traits can cause problems when selecting on a single plant basis in highly heterozygous populations.

4.2.3 Number of segregating families and selections

There have been numerous debates amongst plant breeders concerning the question of how many plants or families should be evaluated, and selected, at each stage in a breeding scheme. Unfortunately, there is no simple recipe to help new breeders, and the questions can only be addressed from an empirical standpoint. Plant breeding is a *numbers game* and the chance of success is often associated with screening many thousands of breeding lines. However, plant breeding programmes should only be as large as the specific breeding group can handle and as resource availability allows. Therefore, it is not productive to grow more lines at any stage in a breeding scheme than can be *effectively and accurately assessed*.

It is often easier to work backwards and ask how many lines can be handled at, say, the advanced yield trial stage in the breeding scheme, and then move backwards to the previous stage and predict how many lines are required at that stage to ensure that the required number are selected, and so on.

Similarly, the number of initial cross combinations that should be used differs markedly in different breeding programmes. Often a large number of crosses need to be screened, as the breeder cannot identify the most productive cross combinations. With experience of specific parents in cross combination and the benefit of ‘cross prediction’ techniques (see Chapter 7), it is possible to reduce the number of crosses screened on a large scale and hence allow breeders to put greater emphasis on cross combinations with the highest probability of producing desirable recombinants and hence cultivars.

4.4 Development of open-pollinated population cultivars

Crops that are generally produced as open-pollinated population include alfalfa, forage legumes, herbage grasses, maize (some), oil palm, perennial ryegrass, red clover, rye and sugar beet.

Development of open-pollinated population cultivars is a process that changes the gene frequency of desirable alleles within a population of mixed genotypes while trying to maintain a high degree of heterozygosity. So it is really the properties of the population that are vital, not individual genotypes (as in pure-line cultivars of self-pollinating crops). Instead of ending with a cultivar for release that is a uniform genotype, the population will be a complex mixture of genotypes, which together give the desired performance.

It is not considered desirable (and it is often very difficult) to develop homozygous or near-homozygous breeding lines or cultivars from these outbreeding species as they suffer severe inbreeding depression, carry deleterious recessive alleles or have strong or partial self-incompatibility systems. There are basically two different types of outbreeding cultivars available: **open-pollinating populations** and **synthetic cultivars**.

4.4.1 Breeding schemes for open-pollinating population cultivars

In open-pollinating populations, selection of desirable cultivars is usually carried out by **mass selection**, **recurrent phenotypic selection** or **selection with progeny testing**. Open-pollinating (outbred or cross-pollinating) cultivars are maintained through open-pollinated populations resulting from random mating.

Mass selection

Mass selection is a very simple breeding scheme that uses natural environmental conditions to alter the allelic frequencies of an open-pollinating population. A new population is created by cross-pollinating two different existing open-pollinating

populations. In this case a representative set of individuals from each population will be taken to be crossed – single plants will not of course be representative of the populations. So it is common (even if mistaken) to select individual plants from each population but to take a reasonably large sample of such plants. How they are crossed depends upon the choice of the breeder, but often Bi-Parental matings (BIPs) are performed where specific parents are selected and hybridized. Alternatively, it is common to collect pollen in bulk and use this to pollinate specifically selected female plants in another population, whilst in many other instances breeders allow random mating or open/cross pollination to occur naturally.

The seed that results from such a set of crosses is grown under field conditions over a number of seasons. The theory of the approach is that genotypes that are adapted to the conditions will predominate and be more productive than those that are '*less fit*'. It is also assumed that crossing will be basically at random and result in a population moving towards equilibrium.

Problems with mass selection are related to partial, or complete lack of, control of the environmental conditions other than by choosing suitable locations for the trials. In some instances it is possible to create disease stress by artificially inoculating susceptible plants with a pathogen to act as spreaders, or by growing very susceptible lines in close proximity to the bulk populations. However, the process is empirical and often subject to unexpected disturbances. It also assumes, as noted earlier, that natural selection is going to be in the direction that the breeder desires – an assumption that is not always justified. It should also be noted that care needs to be exercised in isolating this developing population from other crops of this species, which might happen to be growing within pollination distance.

Recurrent phenotypic selection

Recurrent phenotypic selection tends to be more effective than mass selection. The basic outline of this process is illustrated in Figure 4.5. A population is created by cross-pollination between two (or more) populations to create what is referred to as the **base population**. A large number of plants are grown from the base population and a subsample

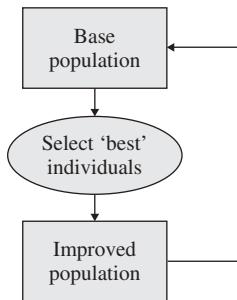


Figure 4.5 Basic recurrent phenotypic selection scheme.

of the most desirable phenotypes is identified and harvested as individual plants. The seeds of these selected plants are grown out, allowed to cross-pollinate at random, and thus to produce seed for a new population – an **improved population**.

This process is repeated a number of times, in other words, it is recurrent. The number of cycles performed will be determined by the desired level of improvement required over the base population, the initial allele frequencies of the base population, and the heritability of the traits of interest in the selection process. Recurrent phenotypic selection has been shown to be effective, but mainly in cases where there is high heritability of the characters being selected for, such as disease and pest resistances. The techniques are not nearly as effective where traits have a lower heritability, such as yield or quality traits.

It is common practice (and a good idea) to retain a sample of the base population so that the genetic changes due to selection can be evaluated in a later season.

Progeny testing

There are actually a variety of possibilities within the main heading of recurrent phenotypic selection, for example half- or full-sib selection with test crossing; and selection from S_1 progeny testing. All the schemes basically involve selecting individuals from within the population, and crossing or selfing these to produce seed. Part of the seed is sown for assessment and part is retained. Once the results of the assessments are available, the remnant seed from the progenies that have been shown to be superior are then sown as a composite population for plants

again to be selected, and so on. At any stage seeds can be taken out for commercial exploitation.

4.4.2 Backcrossing in open-pollinated population cultivar development

Backcrossing is not as commonly used with open-pollinated population cultivar development, unlike pure-line cultivar development, but nevertheless the backcrossing technique is still used. When used, the main difference is that the recurrent and non-recurrent parents in the backcrosses are plant populations rather than homozygous lines. The basic assumption of any backcrossing system is that the technique is unlikely to result in a change in performance of the recurrent parent, other than for the single character being introduced. However, even when an allele has been introduced it is difficult to ensure its even distribution throughout the whole population.

Seed production

In most cases, seed production of open-pollinated cultivars simply involves taking a sample of seed from the population but under increasingly stringent conditions, to avoid the problems noted before, in avoiding contamination or cross-pollination from other populations or cultivars.

4.5 Developing synthetic cultivars

A synthetic cultivar basically gives rise to the same end result as an open-pollinated cultivar (i.e. population improvement), although a synthetic cultivar cannot be propagated by open-pollination without changing the genetic make-up of the population. This has perhaps been a primary reason for the rapid change-over from open-pollinated cultivars to synthetic ones, since it means that farmers need to return to the seed companies for new seed each year. It has been commercial seed companies that have been responsible for breeding almost all synthetic varieties. For example, before 1950 in the USA there were only two alfalfa breeders working in private seed companies while 23 were breeding in the public sector. By 1980 there were 17 public-sector alfalfa breeders, but now there are more than three times (52) the number of private breeders developing synthetic lines.

A synthetic cultivar must be reconstructed from its parental lines or clones. Within the US, maize is almost exclusively grown as hybrid cultivars whereas in many other countries maize crops are grown as synthetics. Synthetics have also been used almost exclusively in the development of alfalfa, forage grass and forage legume cultivars, and have also been used to develop varieties from other crop species (e.g. canola).

The breeding method used for the development of synthetic cultivars is dependent on the ability to develop either homozygous lines from a species or to propagate parental lines clonally. In the case of maize, for example, synthetic cultivars are developed using inbred lines as a three-stage process:

- develop a number of inbred lines;
- progeny-test the inbred lines for general combining ability;
- identify the ‘best’ parents and intercross these to produce the synthetic cultivar.

This process is almost identical to the procedure used for developing hybrid cultivars, and only differs in the last stage where many more parents will be included in a synthetic than in a hybrid cultivar. To avoid repetition, therefore, this section on synthetics will only cover the case of developing synthetic cultivars where inbred line development is not possible (for example, due to high inbreeding depression). So we will deal here with clonal synthetics. The process of developing a synthetic cultivar from clonal lines is illustrated in Figure 4.6.

Following Figure 4.6, clonal selections for use as parents can be added to the nursery *ad infinitum*, on the basis of continued phenotypic recurrent selection of the base population or by selections from those that have been produced by designed cross-pollinations. The second stage involves **clonal evaluation** and is conducted using replicated field trials of asexually reproduced plant units. The aim of the clonal testing is to identify which clonal populations are phenotypically most suited to the environments where they are to be grown. The clonal trials are often grown in two or more locations to include an assessment of environmental stability.

A test cross or polycross technique will then be used to ‘genetically’ test the ‘best’ clones identified from the clonal screen. The aim of this ‘genetic’ test is to determine the general combining ability of each

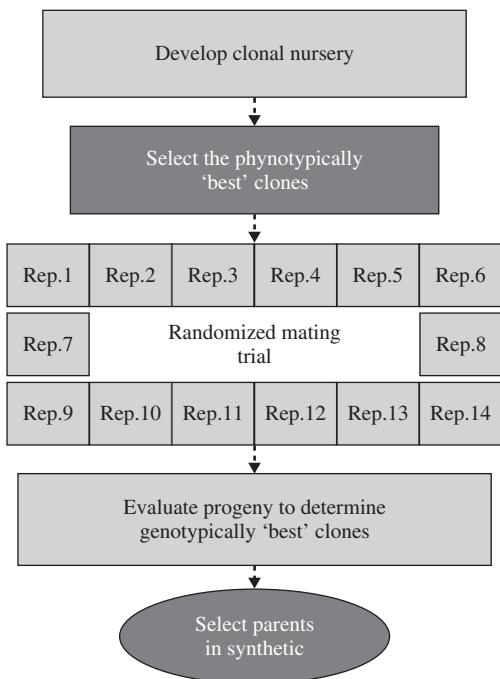


Figure 4.6 Breeding procedure used to develop synthetic cultivars from a clonally reproduced open-pollinated crop species (i.e. alfalfa) and using polycross progeny testing.

clonal line in cross-combinations with other genotypes in the selected group of clones.

If a test cross (often called a '*top cross*') is used, all the selected clones are hybridized to one (or more) test parent. The test parent will have been chosen because it is a desirable cultivar or it may be chosen because of past experience of the individual breeder. The test parent is a heterozygous clone that produces gametes of diverse genotypes. This diversity of gametes produced from the tester will help an assessment of the average ability of each clone to produce superior progeny when combined with alleles from many different individuals. Test cross evaluations are most useful when the variation that is observed within the different progeny is a result of differences between clones under evaluation and not due to only a small sample of genes coming from the test parent.

A polycross does not use a common test parent but rather a number of different parents. It therefore differs from a test cross as the seed progenies to be evaluated result from inter-crossing between

the clones that are under test (i.e. each clone under evaluation is used as female and randomly mated to all, or a good range, of other clonal selections). A polycross, like the test cross, is used to determine the general combining ability of the different potential clonal parents. The seed so produced from a polycross is then tested in randomized field trials. It is essential that the trials are randomized and that the level of replication is high enough to allow the possibility of hybrid seed being from as many other clones as possible.

Seeds from test crossing and polycrossing are grown in **progeny evaluation trials** to evaluate the genotypic worth or determine the general combining ability of each of the clones. Progeny evaluation trials are very similar to any other plot evaluation trial and are best grown in more than a single environment. Progeny evaluations are also often repeated over a number of seasons to obtain more representative evaluations of the likely performance over different years.

Depending on the results from the progeny evaluation trials, clones that show greatest general combining ability will be used as parents for the synthetic cultivar. These clonal parents are mated in a number of combinations to produce experimental synthetics. The number of clones that are selected and the number of parents that are to constitute the synthetic will determine how many possible combinations are possible. If there were only four clones selected, then there would be a total of 11 different synthetic cultivars (that is, the six possible 2-clone combinations, the four 3-clone combinations and the one 4-clone combination). With 6, 8, 10 and 12 parents, the number of possible synthetics would be 57, 247, 1,013 and 4,083, respectively. Therefore it is useful to try to predict the performance of synthetic cultivars without actually producing seed. One formula used is:

$$F_1 - [(F_1 - P)]/n$$

where F_1 is the mean performance of all possible single crosses among n parents, and P is the mean performance of the n parents. It should, however, be noted that there is an assumption of an absence of epistasis (interaction between alleles at different loci) in order to obtain a good estimate of synthetic performance, and so predictions are therefore often far removed from what is actually observed.

Such predictive methods should therefore be used with caution.

4.5.1 Seed production of a synthetic cultivar

The parents used to produce a synthetic cultivar can be intercrossed by hand-pollination to produce the first-generation synthetic (Syn.1). The aim is to cross every parent in the synthetic with all others (i.e. a half diallel cross). This can be difficult with some synthetics (e.g. alfalfa) where the number of parents included is often around 40.

It is therefore more common in situations where many parents are used in a synthetic cultivar to produce Syn.1 seed using a polycross procedure. The selected parents are grown in close proximity in randomized block designs with high replication. The crossing block is obviously grown in isolation for any other source of the crop to avoid cross-contamination. In cases where insect pollinators are necessary to achieve cross-pollination, attempts are made to ensure that these vectors are available in abundance. For example, alfalfa breeders introduce honey bees or leaf-cutter bees to pollinate synthetic lines.

Seed from Syn.1 is open-pollinated to produce Syn.2, which is subsequently open-pollinated to give the Syn.3 population, and so on. The classes of synthetic seed are categorized as breeders' seed, foundation seed and certified seed. In this case breeders' seed would be Syn.1, foundation seed Syn.2, and the earliest certified seed Syn.3.

In summary, the **characteristics of a synthetic cultivar** are:

- Synthetic cultivar species need to have potential to have parental lines, which reproduce from source material (either clonally or as an inbred line), and hence populations with the specific genetic makeup of the synthetic cultivar can be reproducibly produced from these base parents.
- To develop synthetic cultivars, the contributing parental material is tested for combining ability and/or progeny evaluated.
- Pollination of synthetic cultivar species cannot be controlled, and there has to be some natural method of random mating between parents (e.g. a suitable out-crossing mechanism such as a self-incompatibility system or other mechanism to promote cross pollination combined with an appropriate pollinator or wind pollination).

- The source parental material must be maintained for further use.
- Open-pollinated populations have limited life and are then reconstituted from the base parental lines on a cyclic basis.

4.6 Developing hybrid cultivars

Crops that are commonly produced and sold as hybrids include maize, Brussels sprouts, kale, onions, canola, sorghum, sunflower and tomato.

Although attempts have been made to develop hybrid cultivars from almost all annual crop species, it is unfair and incomplete to consider the evolution of hybrids without a brief history of the developments in hybrid corn. At the beginning of the 20th century it became apparent that genetic advances and yield increases achieved by corn breeders were markedly lower than those realized by other small-grain cereal breeders developing wheat and barley cultivars. Indeed, the pedigree selection schemes used by corn breeders, although considered suitable at that time, were essentially not effective. The knowledge that hybrid progeny produced by inter-mating two inbred lines often showed hybrid vigour or heterosis (i.e. transgressive segregation whereby the hybrid progeny produces higher yield than the better parent) suggested that hybrid cultivars could be exploited by corn breeders on an agricultural scale by manually detasseling female parents to produce a population that was entirely composed of F₁ hybrid seed that could be sold for commercial production.

The first suggestion of using controlled crosses was made by W.J. Beal in the late 19th century, based specifically on the concepts of Darwin on inbreeding and outbreeding. These ideas were then refined by G.H. Shull in 1909, on the basis of genetic studies, who put forward the idea of a hybrid cross being produced by first developing a series of inbred, or near-inbred, breeding lines and inter-mating these as inbred × inbred crosses (single cross hybrid) and using the hybrid seed for production. These hybrid progeny were indeed high-yielding and showed a high degree of crop uniformity. His basic concept, however, was not adopted at that time because even the most productive inbred lines had very poor seed yields (most likely due to inbreeding depression) and consequently, hybrid seed production was very expensive.

In the interim the traditional open-pollinated corn cultivars were quickly superseded by double cross hybrids ([Parent A × Parent B] × [Parent C × Parent D]) suggested by D.F. Jones in 1918. Double cross hybrids were not as high yielding or as uniform compared with the single cross hybrids proposed by Shull. However, hybrid seed production was less expensive than single cross hybrids, and as a result double cross hybrids completely dominated US corn production by the 1940s.

Initially all commercial hybrid corn seed was produced by detasseling *female* plants and growing them in close proximity to non-detasseled *males*, and harvesting seed only from the *female* plants. When this method of hybrid corn seed production was most prominent it was estimated that more than 125,000 people were employed in detasseling operations in the US in any growing season. Increased labour costs, combined with developments in using cytoplasmic male sterility (CMS) production systems in the 1960s, rendered detasseling of females in hybrid corn production obsolete for some time, but is now once again the method of choice, although some other biotechnology approaches to this are now being considered.

From the onset of hybrid corn breeding it was realized that there was a limit to production based on the inbred lines available, and a big effort was put into breeding superior inbred lines to use in hybrid combinations. Introduction of efficient and effective CMS hybrid production systems combined with the development of more productive inbred parents led to the return of single cross hybrids, which now dominate corn production in the US and many other developed countries. The relatively high cost of hybrid corn and other hybrid crop seeds does, however, limit the use of hybrid cultivars in many developing countries. Nevertheless, in several South American countries like Argentina, Chile, Paraguay, Uruguay and Brazil, nearly all the maize seed used by farmers is now hybrid because the superior yield they provide largely offsets their higher price.

The rapid increase in popularity and success achieved in hybrid maize could not have occurred without two very important factors. The first is that many countries, including until recently the US, do (or did) not have any Plant Variety Rights legislation or other means that breeders could use to protect proprietary ownership of the cultivars

they bred. There was, therefore, little incentive for private companies to spend time, resources and effort in developing clonal, open-pollinating or inbred cultivars as individual farmers could increase seed stocks themselves, or they could be increased and sold by other seed companies. Hybrid varieties offered the potential for seed/breeding companies to have an in-built economic protection. The developing companies kept all the stocks of the parental lines and only sold hybrid seed to farmers. These hybrids, although uniform at the F₁ stage, would segregate if seed were retained and replanted (i.e. they would be F₂ progenies). Secondly, the introduction of hybrid maize occurred simultaneously with the transition from traditional to intensive technology-based agricultural systems. The new hybrids were indeed higher yielding, but were also more adapted to the increased plant populations, rising soil fertility levels and overall improved crop management of the times.

There are hardly any agricultural crops where hybrid production has not at least been considered, although hybrids are used in still relatively few crop species. The reasons behind this are firstly that not all crops show the same degree of heterosis found in maize, and secondly that it is not feasible in many crop species to find a commercial seed production system that is economically viable in producing commercial hybrid seed. Indeed, if maize had not had separate male and female reproductive organs and hence allowed easy female emasculation through detasseling, hybrid cultivar development might never have been developed, or acceptance would have been delayed by at least 20 years, until cytoplasmic male sterility systems were available.

Hybrid cultivars have been developed, however, in sorghum, onions and other vegetables using a cytoplasmic male sterility (CMS) seed production system; in sugar beet and some *Brassica* crops (mainly Brussels sprouts, kale and canola) using CMS and self-incompatibility to produce hybrid seed; and in tomato and potato using hand emasculation and pollination.

If hybrid cultivars are to be developed from a crop, then the species must:

- show a high degree of heterosis;
- be capable of being managed so as to produce inexpensive hybrid seed;
- not easily be produced uniformly, and have a high premium for crop uniformity.

There are many differing views regarding the exact contribution of hybrids in agriculture. In hybrid maize there would seem little doubt that there have been tremendous advances made. However, this has been the result of much research time and also large financial investments. In addition, it should be noted that the yielding ability of inbred parents in hybrid breeding programmes have been improved just as dramatically as their hybrid products. Most other hybrid crops (with the exception of sorghum) are also outbreeders. Almost all outbreeding crops show degrees of inbreeding depression and, therefore, its counterpart heterosis. In such cases there are strong arguments, certainly in practical terms, for exploiting heterozygosity to produce productive cultivars. This implies that hybrid cultivars can offer an attractive alternative over open-pollinated cultivars or even synthetic lines, although seed production costs will always be a major consideration. In inbreeding crops, hybrid cultivar production is much more difficult to justify on '*biological grounds*'.

Committed '*hybridists*', of whom there are many (especially within commercial seed companies), would argue that:

- Yield heterosis is there for the exploitation.
- Hybrid cultivars are economically attractive to breeding organizations and seed companies.
- Technical problems (usually associated with seed production) are simply challenges to be faced.
- The biological arguments are irrelevant.
- The rate of hybrid cultivar adoption is overwhelmingly rapid in countries where farmers can afford them.

Skeptics (of which there seem to be fewer, or who are less outspoken) argue on the basis of experimental data available to date that:

- There is no good evidence for over-dominance, and so it is definitely possible to develop pure-line, inbred cultivars that are as productive as hybrid lines.
- The economic attractions of seed companies should be weighed against high seed costs for the farmers; that technical skills could be put to better (more productive) use.
- The biological reality is all-important.

These latter sceptics usually, however, accept that in the case of outbreeding species, hybrids give a faster

means of getting yield increases, while in the longer term inbred lines would match them, but in inbreeding crops this differential in speed is not present.

4.6.1 Heterosis

The performance of a hybrid is a function of the genes it receives from both its parents, but can be judged by its phenotypic performance in terms of the amount of heterosis it expresses. Many breeders (and geneticists) believe that the magnitude of heterosis is directly related to the degree of genetic diversity between the two parents. In other words, it is assumed that the more the parents are genetically different, the greater the heterosis will be. To this end, it is common in most hybrid breeding programmes to maintain two or more distinct germplasm sources (**heterotic groups**). Breeding and development is carried out within each source and the different genetic sources are only combined in the actual production of new hybrid cultivars or while testing experimental combinations. For example, maize breeders in the US observed significant heterosis by crossing Iowa Stiff Stalk breeding lines with Lancaster germplasm. Since this discovery, these two different heterotic groups have not been intercrossed to develop new parental lines, but rather have been kept genetically separated for parental development, so crossing and selection has been imposed on each heterotic group separately.

It has proved difficult to clearly and convincingly define the underlying causes of heterosis in crop plants. There are very few instances where heterozygous advantage *per se* has been shown to result from over-dominance. The counterpart to heterosis, inbreeding depression, is generally attributed to the fixation of unfavourable recessive alleles, and so it is argued that heterosis should simply reflect the converse effect. Therefore unfavourable recessive alleles in one line would be masked, in the cross between them, by dominant alleles from the other. If this is all there is to it, then heterosis should be fixable in true breeding lines by the selection of lines with only the favourable alleles. In general it has been found that this simple rationalization does not explain all the observed effects. Thus, the question is whether this breakdown in the explanation is related to a statistical problem of the behaviour of a large number of dominant/recessive alleles, each with small effect; whether the failure

to detect over-dominance is simply a technical failure rather than a lack of biological reality; or whether a more complex explanation needs to be invoked.

4.6.2 Types of hybrid

There are a number of different types of hybrid, apart from the single cross types concentrated on above. The different types of hybrid differ in the number of parents that are used in hybrid seed production. Consider four inbred parents (A, B, C and D) – types of hybrid that could be produced would include:

- Single cross hybrids (e.g. A × B, B × D, etc.)
- Three-way hybrids (e.g. [(A × B) × C, D × (C × A)], etc.)
- Double cross hybrids (e.g. [A × B] × [C × D, C × A] × [B × D], etc.)

Single cross hybrids are genetically uniform, whereas three-way or double cross hybrids are genetically heterogeneous (a three-way cross is less heterogeneous than a double cross). In general single cross hybrids have the highest level of heterozygosity and are more productive than the other two, but on the other hand are most expensive in terms of hybrid seed production.

4.6.3 Breeding system for F₁ hybrid cultivars

The three major steps in producing hybrids are therefore:

- development of inbred lines to be used as parents;
- test cross these lines to identify those that combine well;
- exploit the best single crosses as hybrid cultivars.

The system used to develop hybrid cultivars is illustrated in Figure 4.7. The scheme involves six stages:

- Produce two, or more, segregating populations.
- Develop inbred lines (parents) from each of the two populations.

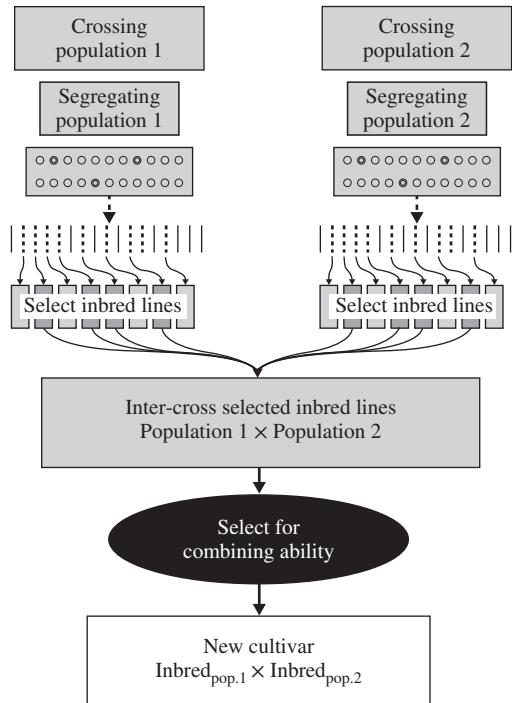


Figure 4.7 Outline of a hybrid breeding scheme.

- Evaluate performance of inbred lines phenotypically.
- Evaluate general combining ability of selected inbred lines.
- Evaluate hybrid cross combinations.
- Increase inbred parental lines, and commercialize hybrid cultivars.

The procedure used to develop inbred parent lines in hybrid cultivar development is similar to that used to breed pure-line cultivars, and the advantages and disadvantages of various approaches are the same. Breeders have used bulk methods, pedigree methods, bulk/pedigree methods, single seed descent and out-of-season extra generations (off-station sites) to achieve homozygosity. Regardless of the breeding approach of choice, these are independently imposed on the different heterotic groups from which inbreds are extracted. One of the most important objectives is to maintain high plant vigour and to ensure that the inbred lines have as high seed productivity as possible. This is not always easy, particularly in species where there is

a high frequency of deleterious recessive alleles present in the segregating populations. Breeders must decide the level of homozygosity that is required. On one hand, the more homozygous (the extreme, of course, being 100% homozygosity) the inbred lines are, then the more uniform will be the resulting hybrid. The more heterozygous inbred lines may, however, be more productive as parents and hence help to reduce the cost of hybrid seed production.

Combining ability (or more relevantly, general combining ability, GCA) is evaluated with the aim of identifying parental lines which will produce productive progeny in a wide range of hybrid cross. Generally, it is not possible to cross all possible parental lines in pairwise combinations, as the number of crosses to be made and evaluated increases exponentially with the increased number of parents. It is therefore more usual to cross each parent under evaluation to a common **test parent** or **tester**. The tester used is common to a set of evaluations, and therefore, general combining ability is determined by comparing the performance of each progeny, assuming that the only difference between the different progenies can be attributed to the different inbred parents. Testers are usually highly developed inbred lines that have proved successful in hybrid combinations in the past. A far better prediction of general combining ability would be achieved if more than one tester were used. This, however, is not common practice, and breeders have tended to prefer to test more inbreds rather than to increase the number of test parents used. As the hybrid programme progresses, it is common, however, to replace older testers with newer ones.

Evaluation of specific combining ability (or actual individual hybrid combination performance) is carried out when the number of parents is reduced to a reasonable level. The number of possible cross combinations differs with the number of parents to be tested. The number of combinations is calculated from:

$[n(n - 1)]/2$	for pairwise crosses
$[n(n - 1)(n - 2)]/2$	for three-way crosses
$[n(n - 1)(n - 2)(n - 3)]/8$	for double crosses

where n is the number of parents to be evaluated. For example, if 20 parents are to be tested then there would be: 20 crosses to a single tester; 190

pairwise crosses, 3,420 three-way crosses and 14,535 double cross combinations possible. It is therefore common to predict the performance of three-way and double crosses from single cross performance rather than actually test them. The three-way cross $[(P_1 \times P_2) \times P_3]$ performance can be predicted from the equation:

$$1/2[(P_1 \times P_3) + (P_2 \times P_3)]$$

where $P_1 \times P_3$ is the performance of the F_1 progeny from the cross between P_1 and P_3 . It is noted that the actual single cross in the hybrid predicted ($P_1 \times P_2$) is not used in the prediction.

To predict the performance of a double cross $[(P_1 \times P_2)(P_3 \times P_4)]$ the following equation is used:

$$1/4[(P_1 \times P_3) + (P_1 \times P_4) + (P_2 \times P_3) + (P_2 \times P_4)]$$

where $P_1 \times P_3$ is the performance of the F_1 progeny from the cross between P_1 and P_3 , and so on. Note again that the two single crosses used in the double cross do not appear in the prediction equation.

The assumptions underlying this will not be discussed here, but we simply note that this is what is carried out quite often in practical breeding.

4.9 Summary

In summary, different crop species are more or less suited for developing different cultivar types. Individual plant breeders have developed specific breeding schemes that best suit their situation according to the type of cultivar being developed, the crop species and the resources (financial and others) that are available.

Irrespective of which exact scheme is used, breeding programmes also differ in the number of individual phenotypes that are evaluated at each stage and in the characters that are used for selection at each of the breeding stages. To address the issue of numbers in the breeding programme, it is necessary to consider the mode of inheritance of the factors of interest, and to have an understanding of the genetics underlying their inheritance.

7

Selection

7.1 Introduction

Selection amongst all living organisms has been going on since life first existed. Natural selection (i.e. evolution) has resulted in the diversity of plant and animal life that exists on Earth today. All selection results in a change of gene frequencies. Throughout evolution, species have been changing, '*more fit*' genotypes have predominated, while those that are less fit with regard to survival or reproduction have become extinct. The aim of plant breeding is to mimic this process, but to direct selection towards increasing the frequency of desirable gene combinations that best suit agricultural systems and increasing consumer needs.

In order to be successful in a selection programme, two criteria need to be satisfied:

- There is variation between plants within the unselected population and the breeders must be able to distinguish between different phenotypes.
- At least some of that variation must be genetic in nature.

Obviously, if a plant breeder cannot distinguish any differences between plants within a population (or different populations), then it will be impossible to select those individuals that appear superior. Secondly, if the variation observed between plants within a population is the result purely of the environmental response of lines, with no genetic component, then there will be no progress made in a

selection scheme as any observed variation will not be inherited.

7.2 What to select and when to select

Having decided that the two criteria above are indeed satisfied, among the first tasks to be addressed by a plant breeder are to decide *what characters are to be selected for*, and *at what stage in the breeding scheme will selection be applied*.

Consider the first question of *what to select for*? To address this, a plant breeder must refer to the **breeding objectives**. These will have been set according to criteria such as:

- the potential market size of the crop;
- the region targeted for propagation;
- the major deficiencies existing within cultivars that are presently available;
- the economic implications of addressing deficiencies such as disease and pest resistance;
- needs of the farmer, such as rapid establishment, early maturity, plant height, resistance to lodging, and harvesting ability;
- the needs of the end-user, including: appearance; uniformity; storability; processing quality; and so on.

Many more factors may need to be included in setting the breeding objectives, and the above list mentions but a few of the more important questions that need to be addressed.

It is not usually possible to select for all the wide range of characters needed for a successful new cultivar in a single season. Plant breeders therefore screen plant populations over several years, sometimes addressing a number of different traits at each evaluation stage. Having decided what characters are to have greatest priority, it is necessary to follow an organized scheme of selection to determine which characters will be addressed at the various stages.

The inheritance of traits will be of great importance in determining not only whether selection is to be carried out, but also the complexity of field and laboratory experimentation needed in order to identify the desirable types.

7.2.1 Qualitative trait selection

Characters, or variation within a character, whose expression show a qualitative form of inheritance can be easily selected for, provided that a suitable screening method is available to determine the presence or absence of the single gene in plants of seed-propagated crops. If the expression of the qualitative character is determined by a recessive allele, then a single round of selection should ensure that all selected plants are fixed for the particular trait. If the desirable allele is completely dominant, several rounds of recurrent selection will be necessary to ensure that the character is genetically fixed in selected plants.

Qualitative characters can often be selected relatively quickly and using very small plots (sometimes even single plant plots) compared with quantitative inherited traits. The ease of selecting for single gene traits has resulted in these characters having high selection intensity in the early generation stages where most genotypes are evaluated, and where it may only be possible to grow small plots because of seed availability.

Selection for such qualitative expression can indeed be a powerful tool in reducing the number of genotypes to be processed, selected and advanced in a plant breeding scheme, although it should never be forgotten that it is often the quantitatively inherited characters that add greatest value to a new cultivar (i.e. yield, quality and durable plant resistance). If early generation selection is to be carried out for single gene traits, then the breeder must be sure that this selection is not having an

adverse effect on the selected populations (i.e. no linkage between advantageous qualitative traits and adverse quantitative traits, or any unwanted non-allelic interactions, or pleiotropic effects).

7.2.2 Quantitative trait selection

Quantitatively inherited characters are usually more difficult to evaluate due to the higher potential for modification of expression by the environment and the larger number of genes involved in their control. Greater experimentation (replication or plot size) is necessary to maximize selection response. As a result, many of the quantitative traits are not positively selected for in the early generation selection stages. Selection for these characters is often delayed until the numbers of different genotypes that require testing are reduced, and where greater amounts of planting material are available for more sophisticated tests. For example, it is common practice in most plant breeding schemes not to select the early generation lines for quality traits that involve either large quantities of produce, or which provide only crude estimates of worth with small samples, or that are expensive.

Obviously any character that is considered of high importance should be selected for at the earliest stages of a plant breeding scheme, where greatest variation will exist among families or populations, but where the trial designs and amount of material make selection effective. Despite the simplicity of this statement, in practice it is often completely ignored.

The characters that are evaluated at different stages of a plant breeding scheme will be discussed in later sections.

7.2.3 Positive and negative selection

Two forms of selection are said to be available to plant breeders: positive and negative selection. It is sometimes difficult to clearly define the difference between the two types (and indeed, some wise and seasoned breeders do not distinguish between them). In simple terms, negative selection is where the very worst plants or families are discarded, while positive selection is where the very best plants or families are selected. In either case, the focus is on the extreme end of the distribution of variation in the trait. Perhaps the simplest

description would be related to the proportion of plants that are selected from a population. If more than 50% of the original population is selected, then this can be considered negative selection. If less than 50% of the population is retained, then this would be positive selection.

7.3 Response to selection

It has already been stated that selection will only be successful if there is sufficient phenotypic variation and if at least some of this variation is genetic in origin. It should be of no surprise, therefore, that the response to selection is related to heritability. Indeed, consider the equation:

$$X_1 - X_{n-1} = R = i\sigma h^2$$

where X_1 is the mean phenotype of the selected genotypes, X_{n-1} is the mean phenotype of the whole population, R is the advance as a result of one round of selection, h^2 is the appropriate heritability (narrow-sense heritability for sexually reproducing species, or broad-sense heritability for clonally or apomictic breeding crops), σ is the phenotypic standard deviation of the whole population, and i is the **intensity of selection**, which is a statistical factor dependent upon the proportion of the population selected. The above equation is probably the most fundamental equation in plant breeding and should be kept in mind.

The intensity of selection (i) is related to the proportion (or percentage) of the population that is selected (k), and takes the values:

Proportion selected (k)	i
0.01	2.665
0.05	2.063
0.10	1.755
0.20	1.400

Although the intensity of selection (i) has been tabulated for a range of different selection rates, in cases where the initial population is large (i.e. greater than 50 genotypes) and the proportion of genotypes selected less than 20%, then the following equation can provide an estimate of i :

$$i = 0.77 + (0.96 \times \log(1/k))$$

From the tabulated values of selection intensities and the estimation equation it can be seen that there is not a linear relationship between higher selection rates (k) and greater response to the selection applied. Retaining 10% of the selected population results in an intensity of selection value of 1.755, while retaining only 1% (i.e. a 10-fold reduction in selections) results in an intensity of selection value of only 1.52 times larger (i.e. $i = 2.665$).

Consider a simple example which is represented diagrammatically in Figure 7.1. Selection is to be carried out on a base population with an average, or mean, of 560 kg yield and with phenotypic standard deviation (σ_p) of 19.0 kg. We assume that from past research it is known that the heritability (h_2) is equal to 0.6 and selection is to be carried out at the 10% level (i.e. $k = 0.1$, $i = 1.755$).

From this we have that:

$$\sigma i = 19.0 \times 1.755 = 33.34$$

From this we can estimate the performance of the selected fraction in the following year, as the response to selection, would be $\sigma i h^2$ and equal to $33.34 \times 0.6 = 20.0$ kg. The mean of the selected plants would therefore be $560\text{kg} + 20\text{kg} = 580\text{kg} =$ the average performance of the top 10% selected lines in the next year.

It should be noted that the phenotypic standard deviation in the selected population must be less than the whole (unselected) population. As it can be assumed that the error variance remains constant, then this must mean that the genetic variance is smaller and the error variance is the same. From this, the heritability between the selected population and further selection years must be less than from the base population in the first selection year.

Therefore, if selection continues, then there would be decreasing response with increasing rounds of selection (Figure 7.2).

Return now to the response equation given above, and recall that the formulae for the broad-sense and narrow-sense heritabilities are σ_g/σ_p and σ_a/σ_p , respectively, where σ_g is the genetic variance component, σ_a is the additive genetic variance and σ_p is the phenotypic variance. From this we can write the average performance of a selected population after selection is:

$$P = X + \sigma i h^2$$

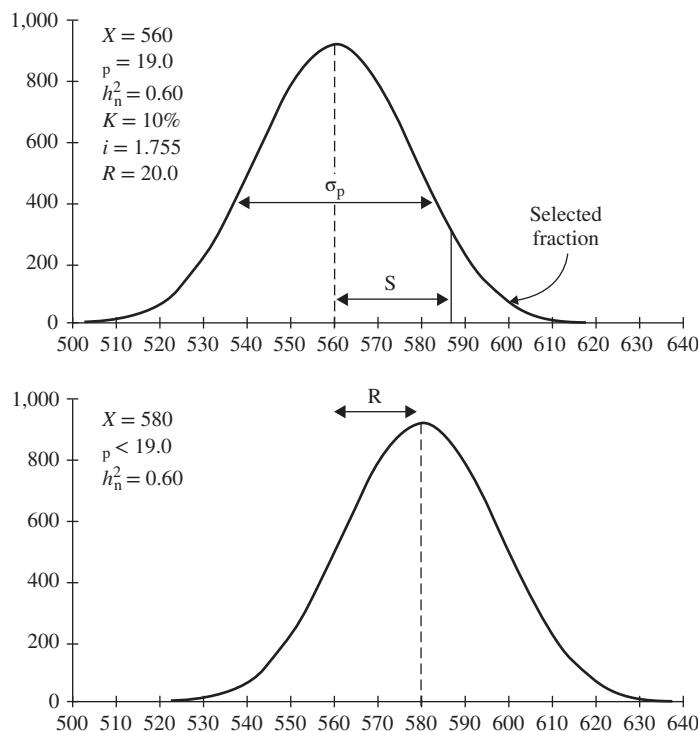


Figure 7.1 Illustration of the response from selection given population parameters from the unselected population (top) to predict the selected population (bottom).

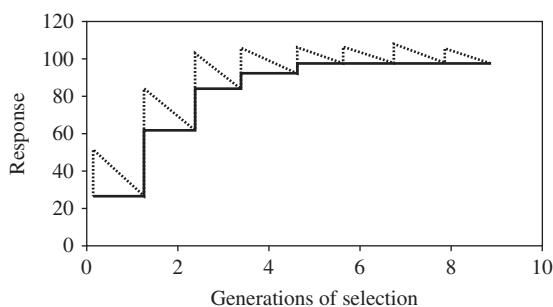


Figure 7.2 Response to selection from successive rounds of selection. The dashed line indicates the phenotypic expression and the solid line represents the genetic gain. Note that greatest gains are from the initial rounds of selection and that after several rounds of selection there is little or no gain.

where X is the average performance of the initial population (i.e. the unselected family mean), i is the selection intensity, h^2 is the heritability and σ is the phenotypic standard deviation between plants in the population.

This means that the very best responses from selection are based on high family means, high selection intensity (although limited increase in return for very high selection), heritability, and the phenotypic variance. From this, breeders should be aiming to:

- Identify highly productive families with high average performance (i.e. high means).
- Maximize heritabilities by minimizing non-genetic errors. This can best be achieved by good experimentation (i.e. selected appropriate approaches for statistical designs and data analysis), improved agronomy, increasing plot sizes and replication levels, and standardized, automated procedures to collect trait data in the field.
- Select as intensely as considered feasible, although remember the efficiency will only increase as a reciprocal beyond 20% selected.
- Choose parents that are genetically diverse for characters that require improvement or change, and hence attempt to increase the phenotypic

variance. Sometimes this could become a hard decision, as the pressures breeders operate under and the limited resources available often encourage them to only cross "best" by "best". Some breeders develop "pre-breeding" populations as a mean to enhance genetic diversity without affecting their probabilities of commercial success, and select and advance lines from these pre-breeding efforts into their mainstream populations.

On the other hand, if a plant breeding programme is not producing the expected response, the same equation can be used to identify possible reasons for the failure.

The close correspondence between heritability and the proportional change in a selected character from one generation to the next, when selection is applied, has already been pointed out. Having considered estimation of narrow-sense heritability, h_n^2 , in some detail earlier, it is now appropriate to return to the issue of estimating heritability.

A third definition of narrow-sense heritability, usually termed the *realized heritability* (h_r^2), is:

$$h_r^2 = R/S$$

where R is the **response to selection** (the same as described above) and S is the **selection differential**. The response to selection is the difference between the mean of the selected genotypes for a particular character and the mean of the population before selection was applied. The selection differential is the average phenotypic superiority for the character in question of the selected genotypes over the whole of the population from which they were selected.

Consider the following example: the average seed yield of an F_3 family is 15 kg. Suppose plants that produced the highest seed yields (with a mean = 20 kg) were selected to be grown in the next generation. What would be the selection differential? Since the selection differential is the average performance of the selected plants over the base population (i.e. all original unselected plants in the family) as a whole, $S = 20 - 15$ or 5 kg seed yield.

Now, if the mean seed yield of the selected progeny in the following year (F_4) was found to be 17.5 kg, what would be the response to selection? Since response to selection is the difference between the mean of the progeny and the mean of

the parental generation before the application of selection, $R = 17.5 - 15$ or 2.5 kg of seed yield.

Finally, the realized heritability (h_r^2) would be given by:

$$h_r^2 = R/S = 2.5/5.0 = 0.5$$

It should be noted that in the above example it is assumed that there are no dramatic year effects. In a practical situation, actual performance from year to year is highly variable and largely unpredictable. This can be taken into account in part by growing a **random sample** of progeny and/or control cultivars the next year. Assuming that the random sample is indeed representative of the whole sample, it will be possible to use this in order to adjust the values and to obtain a direct indication of response to selection.

Similarly, a plant breeder is quite likely to want to know what response might be expected from a given selection differential when the narrow-sense heritability has already been estimated (from the partitioning of phenotypic variances or from offspring-parent regression).

Therefore, if the selection differential (S) applied was 5 kg of seed yield and the narrow-sense heritability had been estimated to be 0.5, the response to selection expected would be:

$$R = h_r^2 \times S = 0.5 \times 5.0 = 2.5 \text{ kg}$$

Thus the average seed yield of the selected progeny might be expected to be 15 + 2.5 or 17.5 kg, and the gain from the selection operation would be 2.5 kg.

7.3.1 Association between traits or years

The degree of association between any two, or a number, of different characters can be examined statistically by the use of **correlation analysis**. As noted earlier, correlation analysis is similar in many ways to simple regression, but in correlations both variables are expected to be subject to error variance, and there is no need to assign one set of values to be the **dependent variable**, while the other is said to be the **independent variable**.

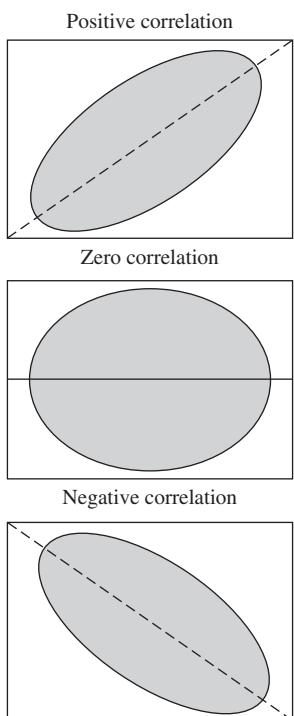


Figure 7.3 Diagrammatic representation of positive correlation (top), zero correlation (middle) and negative correlation (bottom).

Diagrammatically, the association between two variables is shown in Figure 7.3 with positive correlation (top), no correlation (middle) and negative correlation (bottom) indicated. Recall that a correlation coefficient does not in any way establish a cause–effect relationship, only the existence of an association between two variables which may well reflect their association with a third unidentified variable.

As can be seen, if there is strong positive correlation between two traits it will be possible to select individual genotypes that have high expression in both traits. Conversely, if there is strong negative correlation between traits of interest, it may be very difficult to select genotypes with high expression in both characters. The magnitude of the correlation value, in absolute terms, can be associated with underlying physiological processes or even pleiotropy (i.e. the same genes directly control expression in the two characters), or may be a reflection of genetic linkage.

It would seem obvious that there must be some relationship between r , the correlation coefficient, and h^2 , the heritability. If, therefore, characters are recorded on the same set of genotypes grown in two different environments (say locations or years), then the magnitude of the correlation coefficient indicates the relationship between performances in the different environments. Squaring the correlation coefficient (r^2 – “the coefficient of determination”) provides an estimate of the proportion of the total variation between the environments that is explained by the correlation. The total variation between sites can be considered the total phenotypic variation, and as the proportion accounted for by regression must have a genetic base, then a simple relationship exists whereby r^2 is a direct estimate of h^2 .

7.3.2 Heritability and its limitations

In this short but important section, a critical look is taken at the concept of heritability, its uses and misuses.

Four distinct methods of estimating narrow-sense heritabilities have been outlined:

- partitioning of phenotypic variances;
- offspring–parent regression;
- response to selection;
- correlation analyses.

How response to selection can be predicted from a given selection differential, when the narrow-sense heritability is already known from other experiments, has also been covered above. The concept of heritability, and estimates of it, have been of great value to plant breeders and to population geneticists interested in continuously varying characters in natural populations. However, it is very important that the limitations of heritability estimates are realized. These limitations occur on at least three levels:

- There are many technical assumptions inherent in the theory as presented (e.g. that genes assort independently, that alleles segregate independently, and that there is no epistasis). It is possible to allow for many of these complications, but only at the expense of making the theory more complicated.
- An estimate of narrow-sense heritability strictly applies to a *particular* character, in a *particular* population, at a *particular* moment and in a

particular environment. Thus, even for a single character, heritability is not constant. It is obvious that h^2 is particularly vulnerable to changes in environmental variance, changes that can occur at the same place at different times, different places at the same time, or both. Indeed, widely different estimates of h_n^2 for the same character can be found in different populations investigated at the same place and time. Also, it has been seen that additive genetic variance, and hence narrow-sense heritability, generally declines over generations of selection, even in a constant environment. Caution must therefore be exercised in interpreting estimates of h_n^2 if it is not known that every precaution has been taken to expose different populations and/or characters to the same range of environments, and one is interested only in the response in the same, or very similar, environments.

- While means are so-called first degree statistics, variances are second degree. Second degree statistics are usually 'less precise' than first degree statistics. As h_n^2 and h_b^2 (with the exception of h_n^2 from offspring onto mid-parent regression) are based on the ratios of variances, they thus share the weakness of the same lack of precision as all second degree statistics.

2.3 Scientific articles



2.3.1

Hill, W. G. (2010). *Understanding and using quantitative genetic variation.*
Phil. Trans. R. Soc. B. 365:73-85.

Review

Understanding and using quantitative genetic variation

William G. Hill*

Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh,
West Mains Road, Edinburgh EH9 3JT, UK

Quantitative genetics, or the genetics of complex traits, is the study of those characters which are not affected by the action of just a few major genes. Its basis is in statistical models and methodology, albeit based on many strong assumptions. While these are formally unrealistic, methods work. Analyses using dense molecular markers are greatly increasing information about the architecture of these traits, but while some genes of large effect are found, even many dozens of genes do not explain all the variation. Hence, new methods of prediction of merit in breeding programmes are again based on essentially numerical methods, but incorporating genomic information. Long-term selection responses are revealed in laboratory selection experiments, and prospects for continued genetic improvement are high. There is extensive genetic variation in natural populations, but better estimates of covariances among multiple traits and their relation to fitness are needed. Methods based on summary statistics and predictions rather than at the individual gene level seem likely to prevail for some time yet.

Keywords: genetics; animal breeding; quantitative genetics; heritability

1. INTRODUCTION

Traits such as size, obesity or longevity vary greatly among individuals, and have continuously distributed phenotypes that do not show simple Mendelian inheritance.

Quantitative genetics, also referred to as the *genetics of complex traits*, is the study of such characters and is based on a model in which many genes influence the trait and in which non-genetic factors may also be important. The framework can also be used for the analysis of traits such as litter size that take a few discrete values, and of binary characters such as survival to adulthood that have a polygenic basis. The quantitative genetics approach has diverse applications: it is fundamental to an understanding of variation and covariation among relatives in natural and managed populations, of the dynamics of evolutionary change, and of methods for animal and plant improvement and alleviation of complex disease.

On the premise that many genes and the environment act and interact to determine the trait, founders recognized that it would be difficult if not impossible to determine the action of individual trait genes. Statistical methods were invented by Fisher (1918) and Wright (1921), the analysis of variance and path coefficients, respectively, to partition the variation and describe the resemblance between relatives, and such tools and methods developed in

quantitative genetics have had widespread application in disciplines way outwith their original targets.

The models and summary quantities defined by Fisher and Wright have remained at the heart of the subject not least because they provide ways to make predictions of quantities such as the response to artificial and natural selection. Useful parameters include, for example, breeding value (A), which is the expected performance of offspring, and heritability ($h^2 = V_A/V_P$, the ratio of additive genetic variance or variance of breeding value V_A to the overall or phenotypic variance V_P , but widely misunderstood). In view of the assumed complexity of the underlying gene action, involving many loci with unknown effects and interactions, much quantitative genetic analysis has, unashamedly, been at a level of the ‘black box’.

Basic questions range widely: what do the genes do; how do they interact; on what traits does natural selection act; why is there so much genetic variation; and can we expect continued genetic improvement in selection programmes? Ultimately, we want to know at the molecular level not just which genes are involved, whether structural or regulatory, but what specific nucleotide change in each gene or alternatively copy number variant is responsible for the quantitative trait effect, and how the genes are controlled. Much progress is being made in addressing these problems, but many questions remain.

For many decades claims have been made that quantitative genetics was dead or dying but, condescendingly, perhaps still useful until the contents of the black box were revealed, a feat which would be ‘just round the corner’. We are indeed becoming increasingly able to peer inside the box and can ask

*w.g.hill@ed.ac.uk

One contribution of 19 to a Theme Issue ‘Personal perspectives in the life sciences for the Royal Society’s 350th anniversary’.

whether our statistical models of genetic variation in traits are so unrealistic that the edifice may topple. Studies have, however, already revealed almost 50 quantitative trait loci (QTL), many identified to genes, segregating for human height (see later); but these QTL, likely to be individually among the most important, contribute only about 5 per cent of the genetic variation. In view of its complexity, it therefore seems likely that the black box will remain cloudy for a while, even though fed information on, *inter alia*, myriads of genetic markers, levels of gene expressions and trait phenotypes. Statistical methodology which works and is continually developed to incorporate extensive marker and other new data seems likely to remain important for some time yet: better to work with the whole beast rather than try to assemble its parts from inadequate instructions.

I will address some of the background and some of these questions in this personal perspective, which is inevitably uneven in coverage and references, and reflects my interests, biases, knowledge and lacunae. It will focus particularly on animal improvement, an area which has both stimulated many developments in quantitative genetics, and is relevant to the welfare of man. Other recent perspectives and summaries from different viewpoints can be found in, for example, papers by Roff (2007), from the Third International Conference on Quantitative Genetics (2009, *Genetica* **136**, 211–386), and in a *Nature* Insight series (2009, *Nature* **456**, 719–744).

2. THE STATISTICAL FOUNDATIONS OF QUANTITATIVE GENETICS: MODELS, ASSUMPTIONS AND PREDICTIONS

Let us review the standard assumptions in quantitative genetic analysis, address whether they stand up, and if not how much it matters.

(a) Partition of variance components

In the model proposed by Fisher (1918) and developed by Cockerham (1954) and by Kempthorne (1954), variances and covariances among relatives are described in terms of the variances in additive genetic effects or breeding values, V_A , interactions of effects between alleles within loci (dominance, V_D) and among loci (epistasis, V_{AA} , V_{AD} , ...) (Falconer & Mackay 1996; Lynch & Walsh 1998). These partitions are not dependent on numbers of genes or how they interact, but in practice the model is manageable only when the effects are orthogonal, requiring many important assumptions. These include random mating, and hence Hardy-Weinberg equilibrium (*i.e.* no inbred individuals), linkage equilibrium (which requires many generations to achieve for tightly linked genes) and no selection. Gianola & de los Campos (2008) emphasize these, also providing an elegant formalization for the variance-covariance matrix \mathbf{V} of phenotypic values of a group of individuals for a single trait:

$$\mathbf{V} = \mathbf{A}V_A + \mathbf{D}V_D + \mathbf{A}\# \mathbf{A}V_{AA} + \mathbf{A}\# \mathbf{D}V_{AD} + \dots + \mathbf{IV}_E, \quad (2.1)$$

where \mathbf{A} is the numerator relation matrix, or twice kinship (co-ancestry) of individuals, \mathbf{D} defines dominance relationships and V_E the environmental variance. For the epistatic terms, $\#$ denotes element-by-element multiplication, but applies only for unlinked loci. Many more terms may be included, such as maternal genetic effects, and genotype \times environment interaction. The model has unlimited opportunities for complexity. This is a strength, in that it is all-accommodating, and a weakness, in that datasets may be adequate to allow partitioning into only very few components.

(b) Linearity

The regression of offspring phenotype on that of parent for the same or different traits is usually assumed to be linear and, equivalently, so is the regression of response on selection differential. This important assumption holds under multivariate normality of phenotypic and genotypic values and thus the central limit theorem assuming multifactorial inheritance. Some traits, such as litter size or lifespan, are clearly not normally distributed, but adequate transformations can be invoked or departures ignored.

(c) The infinitesimal model

Response to the first generation of selection can be predicted from the breeder's equation $\text{Response} = h^2 \times \text{selection differential}$. Selection changes gene frequencies and hence the genetic variance, so predictions of response in subsequent generations formally require knowing individual gene effects and frequencies. Fisher's 'infinitesimal model', formalized by Bulmer (1980), provides a practical but biologically unrealistic resolution: infinitely many unlinked genes each of infinitesimally small additive effect, so that selection produces negligible changes in gene frequency and variance at each locus. The within-family or Mendelian segregation variance changes only from inbreeding, and the change in between-family variance (the 'Bulmer effect') depends only on the intensity and accuracy of selection practised. Hence the selection response in successive generations can be predicted from estimable base population parameters such as heritability and phenotypic variance, selection practised and inbreeding.

3. DEVELOPMENTS IN STATISTICAL METHODS AND APPLICATIONS

(a) Parameter estimation

Estimates of genetic parameters such as heritability are needed as a basis for description and prediction. Traditional methods such as analysis of variance or regression cannot cope adequately with unbalanced data and the complex pedigrees found outside the laboratory. They have been superseded by more sophisticated methods, often in the context of livestock data (Lynch & Walsh 1998; Sorensen & Gianola 2002), which have been further developed as computing power has increased. An important generalization has been the development of the 'animal model' (*aka* 'individual animal model' or 'individual model') in which the phenotype of each individual is defined in

terms of effects, and the genetic structure is incorporated in the variances and covariances of these effects. For example, a basic model is

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{a} + \mathbf{e}, \quad (3.1)$$

where \mathbf{X} and \mathbf{Z} are design matrices, β is a vector of fixed effects (e.g. years), \mathbf{a} is a vector of random effects (breeding values) and \mathbf{e} is a vector of random errors; and $\text{var}(\mathbf{y}) = \mathbf{Z}\mathbf{A}\mathbf{Z}'\mathbf{V}_A + \mathbf{I}\mathbf{V}_E$ where \mathbf{A} is the additive relationship matrix (equation (2.1)). The model is general and flexible: it can incorporate, albeit with increasing computing needs, other covariance terms such as common environment among full sibs, repeat observations, maternal genetic effects (e.g. birth weight dependent also on dam's genotype as a mother) and multiple traits.

In retrospect, a surprisingly recent development has been in the modelling and analysis of longitudinal traits such as body weight which changes over time. The variances and covariances can be described directly by continuous covariance functions (Kirkpatrick & Heckman 1989) or, equivalently, as parameters of random regression coefficients (Schaeffer & Dekkers 1994).

The generality of the animal model and the fact that most field data (whether humans, livestock or natural populations) are highly unbalanced have created a need for sophisticated and general analytical methods. These use restricted maximum likelihood (REML) or Bayesian principles, facilitated by the availability of specialized computer packages (see reviews by those much involved in their development: Thompson 2008; Sorensen & Gianola 2002). Developments continue, stimulated by the need to deal with non-standard data, e.g. on discrete-valued traits, and to incorporate information on multiple marker genes.

The animal model lends itself to analyses of natural populations, where data are on many traits on a limited number of individuals and the relationship structure is complex. Data are obtained from populations that have been studied long term, such as great tits or red deer, and where births and parentage are recorded or deduced to provide pedigrees (see Kruuk (2004) for exposition and papers in *Proc. R. Soc. B* 275, 593–750, 2008 for examples). Indeed, as genotyping costs fall there are increasing opportunities to expand pedigrees. While relatively simple objectives are to estimate genetic variances and covariances, a broader aim is to use data on breeding success to obtain estimates of the genetic parameters of fitness *per se* (Kruuk *et al.* 2000) and of those characters which determine it, i.e. elements of the selection gradient or partial regression of fitness on each trait. In a natural population, the selection *has* occurred or is currently taking place as a consequence of fitness differences, and a major aim is to infer these selective forces.

The model and methods are flexible but reliable parameter estimation remains a problem and the literature is awash with poor estimates. Few datasets, whether from livestock, laboratory or natural populations, are of sufficient size to obtain useful estimates of many genetic parameters, e.g. there are 30 variances and covariances for four traits when

fitting only additive genetic, sib environment and residual effects, let alone say, dominance, epistasis and maternal genetic effects. We all have our pet ideas as to what are important sources of variation or covariation, and fit models accordingly, but typically many different models can fit almost as well (e.g. full sib common environment and dominance). The animal model can cope with selection and assortative mating, but only if the data on which decisions are based is included (e.g. an analysis on a trait of adults if selection is on any trait of juveniles). Animal breeders encounter many such problems, but they are typically more serious for data from natural populations where datasets may be small, poorly structured and include multiple traits. Some traits associated with fitness, i.e. the selection 'criterion', may not be recorded, and some individuals may die or leave the population before recording. Hadfield (2008), for example, reviews some of these problems and suggests methods for dealing with them.

(b) A new approach: use of high density molecular markers in the partition of genetic variance

Very high density of mapping with multi-locus single nucleotide polymorphism (SNP) chips provides a different method to estimate genetic variances. Pairs of full sibs share 50 per cent of alleles on average, but because linked genomic regions are transmitted, the actual proportion shared varies about expectation, with a s.d. of approximately 4 per cent for humans (Visscher *et al.* 2006). Hence, the genetic variance can be estimated *within* families from the regression of phenotypic similarity of sibs for a trait on the actual proportion of genome shared as determined by SNP identity, and is free of confounding by environmental differences between families or maternal genetic effects (Visscher *et al.* 2006, 2007). Estimates of heritability of human height from this method are about 80 per cent consistent with those from traditional methods. The method can be extended to estimate genotype-sharing among members of non-pedigreed natural populations (including fish), if there is enough money to buy the chips, but relatives providing the most information such as sibs may also share environments.

(c) Prediction of breeding value (or genetic merit) from phenotypic data

Prediction of breeding values is a fundamental component of modern breeding programmes, as those with the highest values should be selected. The major unifying development, Best Linear Unbiased Prediction (BLUP), is due to Henderson (1950, 1984) and incorporates both fixed (environmental) effects and random (genetic) effects in a mixed model (see e.g. Lynch & Walsh 1998; Sorensen & Gianola 2002). As computing power has increased, the animal model (equation (3.1)) is now used, enabling simultaneous prediction of breeding values for all traits of individuals differing in age, location, numbers of records and numbers of relatives. As all selection candidates can be compared at frequent

intervals, with overlapping generations it is possible to cull and select continuously.

BLUP is best in the sense of minimum variance among linear predictors, but only if population parameters are well estimated. It is unbiased in that, as more data are accumulated, the predicted breeding values approach the true values; and while it allows for selection, requires the important but often unachievable proviso that *all* information on all traits on which selection is practised is included in the data. Further, if any selection is practised, the infinitesimal model assumption is implicit (but often forgotten) in the use of the relationship matrix \mathbf{A} to quantify variances and covariances across generations.

4. THE STATISTICS IN PRACTICE: INVESTIGATING AND INFORMING THE ASSUMPTIONS

Many major assumptions are made in the applications of quantitative genetics, but the issue is not the formal correctness of models used, rather the extent to which they work reasonably well. There is not space for a full review, but more discussion and examples are given elsewhere (e.g. Falconer & Mackay 1996; Lynch & Walsh 1998; Walsh & Lynch 2009). We first consider quantitative data at the whole trait level before considering information from studies of QTL and genes.

A major problem is to obtain data of adequate structure and quantity. For example, in the infinitesimal model all genetic variation is assumed to be additive. In random mating populations it is, however, usually impossible to estimate epistatic variances with any precision because the coefficients are very small and highly correlated with those of non-epistatic components (e.g. \mathbf{A} and $\mathbf{A} \# \mathbf{A}$ matrices in equation (2.1)). These in turn may be confounded with other parameters, such as genetic maternal effects to explain why, say, a daughter-dam correlation exceeds twice that of half sibs in the absence of epistasis. Linkage disequilibrium (LD) is patently present, but that owing to close linkage is assumed absent in the infinitesimal model. The orthogonality assumptions in equation (2.1) may not hold, but how should that be tested? Hence, much of the evidence based on quantitative information is unsatisfactory in being so inconclusive, for example in failing to reject even the infinitesimal model as the following examples show.

In a classical study Clayton *et al.* (1957) found good agreement between heritability estimates from different sources and with predictions of selection response. Sheridan (1988), however, showed that there are frequently wide differences between selection responses predicted from base population parameters and those actually realized, but his analysis failed to take into sufficient account the sampling errors of the predictions or the responses (Walsh & Lynch 2009, ch. 14). It is a common observation that regressions of progeny on parent phenotype are roughly linear, but in detailed studies failures can be found (e.g. Gimelfarb & Willis 1994). Frankham (1990) has shown that selection responses for fitness-associated traits are generally asymmetric, faster

down than up, as might be anticipated with a previous selection plateau. We have tried direct application of the infinitesimal model predictions using REML/BLUP to mouse selection experiments, but with inconsistent results: for example a rather poor fit for feed intake in one line (Meyer & Hill 1991), but an excellent fit despite a four-fold change in body fatness in another (Martinez *et al.* 2000). Under the infinitesimal model, the pattern of response in finite populations is predictable from base population parameters. Using data summarized by Weber (2004) on responses at generation 50 relative to those in the first generation, we showed that ‘realistic’ models based on distributions of gene effects, including some of the large effects, provided a good fit to the data; but an infinitesimal model (including mutation) fitted almost as well (Zhang & Hill 2005a). Perhaps, this robustness is unsurprising: Barton & de Vladar (2009) show that the population dynamics can be modelled well using approaches from statistical mechanics, where the population is described solely in terms of stationary distributions of gene frequencies and continued response is insensitive to the details of the genetic architecture.

I am not aware of any ‘experiment’ in which a combination of say REML and subsequent BLUP predictions has been formally tested *in vivo*. Hence, let us take a pragmatic view: if something works in practice is that not sufficient even if the theoretical foundations are generally unsubstantiated? For over 30 years BLUP and related methodology have dominated genetic evaluation of dairy cattle, and models have become increasingly complex. The spectacular genetic improvement achieved is illustrated in figure 1 and is in accord with the infinitesimal model and BLUP prediction.

So while the genetic models adopted may be very crude, their generally satisfactory behaviour explains why many scientists and practitioners applying quantitative genetic principles do not lose much sleep over model assumptions. We are, however, getting new kinds of information from studies at the individual QTL or gene level which should inform, improve, or in due course may replace the classical models and methods. The path from primary gene effect to phenotype may be complex, however; increasingly so as more genes are involved. Even when the genetic lesion is known, using that information to effect a ‘cure’ may be far from straightforward, as the work with cystic fibrosis shows (Pearson 2009).

5. NUMBERS OF GENES, THEIR EFFECTS, THEIR ACTIONS AND INTERACTIONS

Since the time molecular markers became available, extensive studies have been undertaken on analyses to identify QTL and, on occasion, the actual gene or nucleotide (QTN). Indeed, this has been the big quantitative genetics industry of the last two decades. The basic methods are to use associations generated by linkage or LD between marker genes and the trait to locate QTL or to identify and locate mutations having a phenotypic effect and a molecular signal, such as transposable elements. Linkage studies

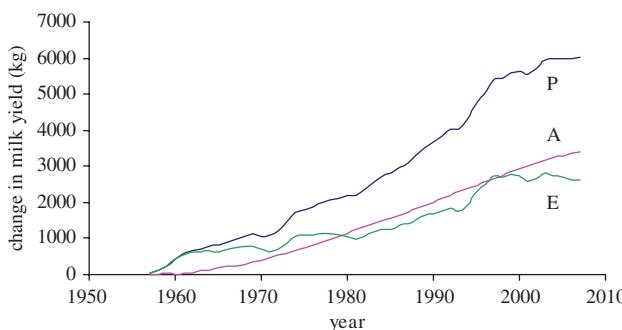


Figure 1. Changes in milk yields of US Holstein cows: phenotypic mean yields (P), mean breeding values (A) and environmental effects ($E = A - P$) derived from USDA data. Results are given relative to 1957, when the mean yield was 5859 kg. (Adapted from <http://apipl.arsusda.gov/eval/summary/trend.cfm>).

(Lander & Botstein 1989; Haley & Knott 1992) have been conducted in designed studies using crosses of inbred lines or, for example, breeds, and family studies in humans. In view of the few recombinants generated in any region of the genome, the linkage studies are usually unable to provide precise location of QTL in the genome even when many markers are available, and in many cases have not been conducted on a sufficient scale. The availability now of dense SNP maps enables and requires data for analysis in which many generations of recombination between markers and QTL may have occurred to enable fine-scale mapping. In the laboratory, recombinant-inbred lines (RIL) have been developed from crosses of multiple inbred lines to introduce much initial diversity (Chesler *et al.* 2008) and multi-line segregating populations established from inbred crosses have been generated (Valdar *et al.* 2006). As for inbred line crosses, the RIL have the further benefit that animals of identical genotype can be generated and many traits studied in relevant specialized laboratories to make the best use of development time and costs. Association mapping using LD enables high-precision mapping in humans, livestock and natural populations, but requires large datasets and high-density SNP marker panels to be effective. Further, it enables inferences to be drawn about frequencies and effects of genes actually segregating in populations. In view of the large resources needed, it is not surprising that most of the information so far generated from association mapping is on human disease; but these and other traits recorded in such studies, for example height, are already providing an important source of information for all quantitative geneticists.

There is an extensive literature on the basic methodology of QTL mapping (e.g. Lynch & Walsh 1998; Weller 2009) and, for example, Mackay *et al.* (2009) summarize both methodology and achievements. There are many statistical problems involved, even in the most basic QTL mapping studies. Not least is the problem of trade off between power of detection and type-I error, with very extreme significance thresholds having to be set when searching over all the many possible sites in the genome. Hence, the QTL most likely to be found are those of largest

effect; very many are likely to be missed; and the estimated effects of those detected are likely to be biased upwards and their position poorly located.

(a) Some examples

Rather than attempt to review or even summarize the field, I shall just give some examples of the results from the use of different techniques, roughly in descending order of precision, that both provide information and generate questions.

In a summary of the analysis of around 600 P -element insert lines in *Drosophila melanogaster*, a method permitting precise location, Mackay (2009) found that about 17 per cent of the insertions affected sensitivity to the inebriating effects of alcohol (even *Drosophila* have an excuse) and 34 per cent affected locomotor behaviour to a stimulus; and she noted that similar screens have found 22 per cent of insertions affecting abdominal and 23 per cent affecting sternopleural bristle number. Some have large effects, however. In view of the fact that such a high proportion of sites are targets, it is not surprising that there is extensive pleiotropy. Mackay also notes that many show epistatic effects. Similarly, for a range of behavioural traits in mice, in a study of over 200 gene knockout lines, 19 per cent showed abnormal open-field activity (Flint & Mott 2008).

Heterogeneous stocks established by crossing inbred mouse lines can allow fine-scale mapping. In an analysis of 97 traits, including body weight and many biochemical variants, of 843 QTL detected and mapped to within 3 Mb, only 10 individually contributed more than 10 per cent of the variance for any trait and none over 3.5 per cent for body weight or length (Valdar *et al.* 2006). A plot of the distribution of QTL contributions to variance shows a peak at about 2 per cent, though it is likely this is, in effect, a truncated exponential-shaped distribution, as smaller ones are non-significant and missed. In principle, such distributions (obtained also in other studies) can be extended to smaller effects, but some prior distribution must be assumed.

The association studies undertaken with combined samples of 10 000 or more humans are revealing a substantial number of QTL that have been cross-validated and in many cases identified to specific genes. Visscher (2008) and Weedon & Frayling (2008) provide summaries. Some 44 independent variants that affect stature, none of which are rare in the population, have been mapped; but none individually explain over 0.5 per cent of the phenotypic variance. The heritability of the trait is about 80 per cent, and overall only about 5 per cent of the variance has so far been accounted for. None of the variants show evidence of departure from additive gene action, i.e. dominance or epistasis, and the difference between homozygotes is about 0.8 cm (or a little over 0.1 phenotypic s.d.). Although the causal genes have not yet been proven, there is a strong candidate in over half the cases. Of these, many are components of signalling pathways known to be important in skeletal growth and development, demonstrated for example by gene knock-outs in mice (Weedon & Frayling 2008).

For cattle, in July 2009 there were 1375 QTL curated into the database (the cattleQTLdb, <http://www.animalgenome.org/>), and likely others were discovered by companies but not entered. These were from 83 publications and represented 109 different traits (but many have pleiotropic effect), representing a major effort and expenditure. The number of animals involved in each analysis are far smaller than in the association studies in humans, although data are used from segregation within individual sires who have progeny-tested sons with accurate estimates of breeding value. As only few of the QTL have been finely mapped, there is uncertainty about which of those mapped in different studies to similar genomic regions are the same or different genetic lesions, and how many are false-positives. In a few cases in livestock the actual genes, all having large effect, have been identified and sequenced. Some were already known as major genes, such as double muscling in cattle, for which the myostatin gene has been identified as causative, and others were initially discovered in mapping studies, for example DGAT, which influences milk composition of dairy cattle (see for example Hu *et al.* 2009 for more examples and references). It is not clear yet if there is any general pattern about what genes will be found to act, but clearly some of the large effects are segregating.

6. CONCLUSIONS ON ARCHITECTURE AND THE 'MISSING' HERITABILITY

The different kinds of analysis are revealing that many loci contribute to quantitative genetic variation. This finding is no surprise to quantitative geneticists because the polygenic and specifically infinitesimal models of quantitative genetics have been shown to work so well in prediction, in distributions and in describing long-term selection response, and the more optimistic expectations in early days of QTL mapping of finding a few regions contributing most of the variation was unrealistic. Indeed predictions made by, for example, Robertson (1967) of contributions of increasingly many genes of increasing small effect have generally been borne out.

While the most reasonable hypothesis to explain why most of the genetic variation in human height is not accounted for by the 50 or so loci contributing most is that there are many more, perhaps thousands, of small effect and more extreme frequency, concern has been expressed about the 'missing heritability' and various hypotheses proposed (Maher 2008). One is that previous estimates of the heritability are biased by environmental correlations, another that various interactions are responsible. But both are refuted by the within-family analysis of Visscher *et al.* (2007, see above) which gives similar estimates of heritability, shows no evidence of interactions across chromosomes, and a distribution of variance contributed roughly proportional to chromosome length. Rare variants including rare copy-number variants could explain some of the variation, as these would contribute to the estimates of within-family variance, but their effects would be hard to detect with the current

resolution of SNP chips. Transient epigenetic effects could contribute to heritability estimates from close relatives (Slatkin 2009), but cannot be a predominant feature as they would not contribute to long-term selection responses.

Perhaps human height is exceptional, for it has a very high heritability and near additivity of variance. Recent association studies on other traits are, however, also revealing many regions of the genome associated with disease risk: almost 20 for type II diabetes (Donnelly 2008), and for schizophrenia, also highly heritable, as significance thresholds attached to individual markers detected in one subset of data were reduced, increasingly more risk could be accounted for in independent sets of cases (Purcell *et al.* 2009). Therefore, the current sample sizes available for genome-wide associated studies are not sufficiently powered to detect the majority of the associated variants.

Neutral genes have an expected U-shaped frequency distribution, $f(p) \propto [p(1-p)]^{-1}$, under rare mutation drift balance (Wright 1931), such that if they are additive the variance is contributed uniformly across gene frequencies. Mutant genes under natural selection, either because they have pleiotropic effects on fitness or are subject to stabilizing selection, show a distribution more heavily weighted to extreme frequencies (Wright 1931; Zhang & Hill 2005a), such that the variance contributed may also be U-shaped. Such loci are hard to detect in association studies even if they have large effect, partly because they contribute little variance and partly because SNP markers that have intermediate frequencies cannot have high correlation in frequency (r^2) with a rare QTL. The hypothesis that most of the missing variation is associated with extreme frequencies is not, however, supported by the schizophrenia study (Purcell *et al.* 2009).

Another important property to be revealed from such studies is the magnitude of pleiotropic effects of genes on other traits. In view of the large number of height genes already revealed but counting for 5 per cent or less of the variance overall, there must be so many genes affecting it overall that pleiotropy for other traits must be widespread. This accords with the findings of Mackay (2009, see above) from mutagenesis studies. In contrast, in an extensive linkage-based line analysis of mouse skeletal measurements, Wagner *et al.* (2008) concluded that pleiotropic effects were rare. But they set significance thresholds at the same high values for detecting pleiotropic effects as for initial detection, such that even a QTL with exactly the same large effect on each trait would be significant for only a few.

A quite different source of evidence on the role of multiple genes comes from the analysis by Laurie *et al.* (2004) of the Illinois maize experiment (see figure 2, discussed further later) with selection for high and low oil content in the kernel. From a line cross made at generation 70 and maintained by random mating for 10 generations to reduce LD, they estimated that about 50 QTL contributed to the response, none exceeding about 0.3 per cent oil to the line divergence of 17 per cent oil. Furthermore,

the QTL acted essentially additively with each other and similarly in pure lines and crosses.

Some of the data provide puzzles about how genes act on quantitative traits which no doubt will take some unravelling. In contrast to the extensive QTL association-based mapping studies in humans showing additive gene action and the useful properties of the infinitesimal model, some studies in livestock, plants and laboratory animals have revealed dominance and epistatic interactions (e.g. Carlborg & Haley 2004; Mackay 2009). Can these data be squared?

Interactions, which are second-order effects, are likely to be tiny and very hard to detect if the main effects are already small. Further, unless all the interacting genes are at intermediate frequencies, they are expected to contribute mostly additive variance simply on statistical grounds (Crow 2008; Hill *et al.* 2008). In inbred line cross or mutagenesis experiments, those loci of large effect that can generate most interactions are more likely to be observed than in outbred natural populations, where their heterozygosity is low if they have any deleterious effect on fitness. So we should not necessarily infer wholly additive effects from additive variance.

In view of geneticists' success in unravelling the control of developmental pattern, it would seem straightforward to figure out how the overall size of the organism is controlled. But we now know that many more than 50 genes affect stature, and arguably all 20 000 genes affect all quantitative traits, together with other controlling factors in the genome. So how in the body is the phenotype determined? One can see a role for systems biology, but I am pessimistic about the rate at which the systems will be disentangled: understanding models for connecting tens of interacting genes may be feasible, but not for 1000. So while we will get a lot more information, I do not believe the essentially statistical approach, enhanced by the use of genomic information to mark genomic regions, is on its last legs.

7. MAINTENANCE OF VARIATION IN QUANTITATIVE TRAITS

Let us turn now from considerations of how quantitative traits are determined to trying to explain why they are so variable in natural and derived populations.

The magnitude of variances and heritability is a property of that population and environment, as it depends on the frequency and effects of the segregating genes, but for the same trait or type of trait they tend to be roughly similar, not just across populations but even across species. Heritabilities (h^2) tend to be highest for conformation traits and mature size, typically 50 per cent or more, and lowest for fitness-associated traits such as fertility (e.g. Mousseau & Roff 1987; Falconer & Mackay 1996; Lynch & Walsh 1998). Conversely, the 'evolvability' or genetic coefficient of variation ($CV_A = h \times CV$), is typically higher for fitness-associated than conformation traits (Houle 1992). Estimates of variance for fitness itself are hard to obtain, but the laboratory-based estimate of V_G for log fitness is 17 per cent (Fowler *et al.* 1997) and although life-history traits in natural

populations show clear evidence of genetic variance, their heritabilities are low (e.g. Kruuk *et al.* 2000). We still seek adequate explanation of what determines levels of quantitative genetic variation, why there is some consistency across populations and species, and why there is so much for fitness-associated traits despite the inference from Fisher's fundamental theorem of its loss by natural selection.

(a) Mutation and genetic variation

Estimates of the amount of genetic variance contributed by mutation, generally expressed as the 'mutational heritability', V_m/V_E , show a surprisingly narrow range over many traits and species, centred about 0.1 per cent (Keightley & Halligan 2009), and equivalent to an increment in CV_A of ca 0.3 per cent for a trait with a CV of 10 per cent. If all genes were neutral with respect to fitness, $N_e = 250$ would maintain a heritability of one-third at the equilibrium $V_A = 2N_eV_m$, but unsurprisingly this close relationship between heritability and population size is not found as many mutations are deleterious.

The most studied model for natural selection acting on the trait directly is stabilizing selection, i.e. intermediates fittest. Under this model, genes of large effect contribute more variance when segregating, but have lower expected heterozygosity, and so the predicted variance maintained is proportional to the total mutation rate to trait genes and inverse of the strength of selection (i.e. curvature of fitness surface). If few loci are assumed to affect the trait and typical estimates of the strength of selection are assumed, the predicted variance is much lower than that observed (Turelli 1984). But even if hundreds of genes affect any single trait, the model is not rescued because mutants are likely to have pleiotropic effects on many traits and overall be under stronger selection than on the target trait alone. The finding of segregating genes at intermediate frequencies affecting human height, for example, indicates that selection pressures are weak, and so both population size and selection set upper bounds to the variance maintained (Bürger 2000). Disentangling selection on multiple traits is difficult or impossible; indeed, there is little evidence for stabilizing selection and as much for its converse, disruptive selection, in the summary of published results by Kingsolver *et al.* (2001).

An alternative model is to assume that the selection does not act on the target trait directly but is through pleiotropic effects of the mutant (Keightley & Hill 1990). This does not, however, resolve the dependency of V_G on population size, nor explain the constancy of trait means. Various aspects of the fit to the data are enhanced both by assuming that the mutants are (nearly) additive for the trait but (partially) recessive for fitness, and worsened by assuming that there are substantial pleiotropic effects on other traits and overall fitness (Zhang & Hill 2005b).

There are a plethora of other models, invoking spatial and/or temporal variation in the environment, competition for resources and (even) heterozygote superiority, but none are clear winners. Johnson &

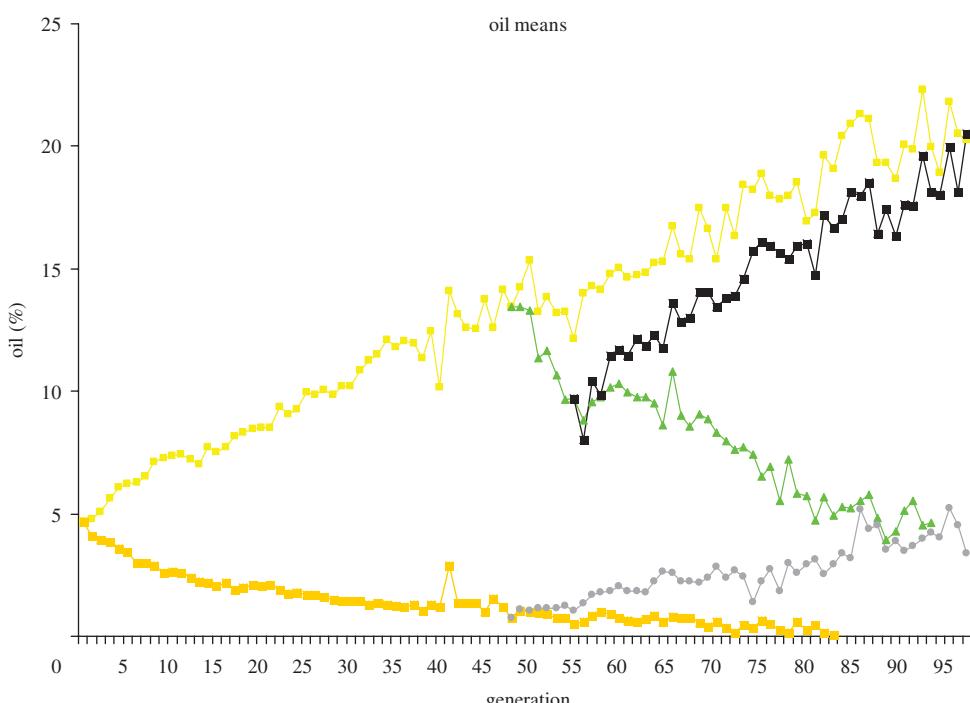


Figure 2. Responses to selection for oil content in maize in the Illinois selection lines. Line designations: IHO (light yellow square), continued selection for high oil, ILO (dark yellow square), for low oil; RHO (green triangle), RLO (white circle), reverse selection; SHO (black square), re-reversed selection. (Adapted from Dudley & Lambert 2004).

Barton (2005, p. 1419) put it well: ‘We are in the somewhat embarrassing position of observing some remarkably robust patterns, that are consistent across traits and species, and yet seeing no compelling explanation for them.’ It is not yet clear how the new genomic data will help, in view of the many genes identified for height, for example. Indeed, the theory requires some rethink to account for the large number of small effects and pleiotropy, and put to best use the new genomic, proteomic and other data that become available.

Further, how the level of phenotypic or environmental variance and hence h^2 are determined has been less studied and is even less well-understood than that of V_G . Evolution of V_E requires genetic variation of phenotype given genotype, for which there is strong evidence in *Drosophila* (Mackay & Lyman 2005) and in livestock populations (e.g. Sorensen & Waagepetersen 2003). Under stabilizing selection genotypes expressing less variable phenotypes are fitter, leading to evolution to reduce V_E . We have suggested two models that would lead to a balance: an ‘engineering’ cost in resources to obtain and maintain homogeneity; and/or most mutations disrupt the phenotype and tend to increase V_E (Zhang & Hill 2008), for which there is some evidence (Baer 2008).

8. LOOKING TO CONTINUED SELECTION RESPONSE AND GENETIC IMPROVEMENT

Let us consider how genomic and individual QTL or gene information can be used in improvement programmes, and what are the opportunities for continued response using straightforward selection

on the quantitative trait and incorporating other technology?

(a) Using individual quantitative trait loci

There has been extensive theoretical analysis and simulation to develop methods for using individual QTL in plant and livestock breeding programmes by marker-assisted introgression of a QTL from another population or by marker-assisted selection to increase frequency of a segregating gene in the population (e.g. Weller 2009). Clearly, its effectiveness depends on the real effect of the QTL, the relation between the predicted and the real effect, the closeness of available markers to the QTL (obviously best if the actual gene is known), and on its frequency in the population; and its impact will be the greatest when phenotypes are absent (e.g. sex-limited traits) or of low heritability.

Much effort has been expended on QTL detection and on theoretical analysis of how best to incorporate them in improvement programmes. We have much less information on actual effectiveness because much is within commercial companies and conventional selection on continuous traits has continued alongside. In two recent reviews on applications in plant breeding, Collard & Mackill (2008) and Hospital (2008) suggest that the great opportunities have not yet been fully realized. In a comprehensive review on work in livestock, Dekkers (2004) concluded guardedly that ‘The current attitude to marker assisted selection is one of cautious optimism’. I consider that the returns from the extensive R&D on QTL identification in livestock have been low, both because selection responses have been high from conventional

selection (e.g. figures 1 and 2) and because estimates of QTL effects and genome location are poor for the lowly heritable traits that are hardest to improve by selection.

(b) Using genomic selection

The availability of marker panels of thousands of SNPs does, in contrast, appear to be bringing in a real paradigm shift following the pioneering study of Meuwissen *et al.* (2001), and seems likely to be less of a false dawn than the use of individual QTL (or indeed of transgenics). The objective is to predict the breeding values of candidates for selection not by identifying just a few QTL of large effect but, by densely marking the whole genome, to incorporate most variants using historical LD in the population. This information is used to assess sharing of genomes of relatives and to weigh the marker genotypes according to the phenotypic effects associated with each region and the imprecision of estimation of these effects. In view of the close linkage, the LD between markers and genes is unlikely to change rapidly over generations, such that it may be possible to use much less dense marker panels after the initial evaluation (Habier *et al.* 2009).

Development of methodology continues, particularly of the statistical methods required to undertake the BLUP predictions. One approach is to replace the expected relationship matrix \mathbf{A} (equation (2.1)) by the realized relationship matrix as assessed using high-density markers (Hayes *et al.* 2009). Another is to more overtly make use of possible differences among genomic regions in contribution of variation in the trait, but if it is assumed that the variance in the trait associated with each SNP is sampled from the same normal distribution, the methods are equivalent (Goddard 2009; Hayes *et al.* in press) and can be used by extension of BLUP methodology, ‘genome-wide BLUP (GWBLUP)’. Under the assumption that a limited proportion of the genome contributes most of the variation, selective procedures have been developed, initially by Meuwissen *et al.* (2001), to identify these regions using a Bayesian analysis with some assumed prior distribution of the number and effects of QTL; but choice of the prior remains controversial.

The methods have widespread potential applications in breeding programmes and can incorporate any number of traits and availability of phenotypic records. Benefits are most obvious in the improvement of sex-limited traits, such as milk or egg production, where young sires have to be selected on the basis of their ancestors’ and female sibs’ records, and all full brothers have the same predicted breeding value. With the genomic information, the Mendelian sampling contribution to each individual son can be predicted. While more research is clearly needed to optimize methodology, genomic selection is now being introduced in widespread commercial practice, a rapid uptake of ideas first published less than 8 years ago (Meuwissen *et al.* 2001).

The USDA provided the first set of genomic breeding values predicted by GWBLUP for bulls in the USA

in January 2009. By making BV predictions for bulls using only data available on their sires, comparisons between predictions with and without the use of genomic information could be made using these bulls’ actual progeny performance. For milk yield, for example, the predicted and observed accuracies using just ancestral phenotypic data were 0.35 and 0.32, and by incorporating the genotypic data, the respective figures increased to 0.69 and 0.56 or 0.58 according to whether differential weights were given to different genomic regions (van Raden *et al.* 2009). In the context of dairy cattle improvement, such near doubling in the accuracy of selection is spectacular. Other studies have shown increases in accuracy, but not all as high as expected, for example on a pedigree population of mice (Legarra *et al.* 2008). Although these need to be understood, for example in terms of numbers of SNPs, the prospects are high, but we await outcomes.

The ideas of genomic selection can be applied to predict disease risk in humans or among selection candidates in livestock, using information on genome sharing with close or more distant relatives (Wray *et al.* 2007). The basic assumption is that many loci contribute to risk, as borne out by analysis at least for schizophrenia (Purcell *et al.* 2009). Perhaps this way, personal genotyping will yield benefits if analysis is put in the hands of those understanding the statistical methodology and its limitations.

Genomics is not the only ‘omics that may provide important information on quantitative traits, and there are alternative ways to use genomic data, such as non-parametric methods (Gianola & de los Campos 2008) that do not use all the Mendelian information. Major developments in the technologies and their use will surely be made. For example, gene expression arrays yield data on thousands more ‘traits’, each individually susceptible to quantitative genetic analysis, and some may well be relevant to particular objective traits. Again some caution is required: physiological predictors of performance, e.g. use of hormone levels, have been much mooted but produced little of practical benefit in livestock improvement. So, overall, it is a question of ‘watch this space’: the extensive new data should be of value for incorporation as ‘markers’ and also new understanding of the biology will be important in its own right and should lead to more effective breeding programmes.

(c) Maintaining selection response, genetic improvement and evolutionary opportunities

We see the striking changes that have been produced in quantitative traits by selection, for example among breeds of dogs in body weight and behaviour, and in the productivity of modern livestock and crops. Can we expect continued change?

The Illinois maize selection for high and low content of oil in the kernel has been continued since 1896. The low lines have reached a plateau (almost 0% in the low oil line, and presumably at the minimum for seed viability in the low protein line), but the upward lines have continued responding for 100 generations (i.e. years, figure 2). Large and continuing

Table 1. Comparison of weight at eight weeks and body composition in two trials, the first of 1957 control and 1991 commercial and the second of 1957 control and 2001 commercial broilers reared on a diet using typical specifications of that year. (The difference D_1 denotes changes between 1957 and 1991 and D_2 between 1957 and 2001, and $D_2 - D_1$ is the estimated change between 1991 and 2001. (Adapted from Havenstein *et al.* 1994, 2003; G. A. Havenstein 2008, personal communication))

year of population	1991 trial			2001 trial			difference $D_2 - D_1$
	1991	1957	D_1	2001	1957	D_2	
body weight (kg)	3.11	0.79	2.32	3.95	0.81	3.14	0.82
carcass weight (kg)	2.07	0.50	1.51	2.81	0.48	2.33	0.82
carcass yield (%)	69.7	61.2	8.5	74.4	60.8	13.6	5.1
breast yield (%)	15.7	11.8	3.9	21.3	11.4	9.9	6.0
carcass fat (%)	15.3	9.4	5.9	15.9	10.6	5.3	-0.6

responses have been seen in other laboratory experiments spanning 100 or more generations (Hill & Bünger 2004). Genetic change in crop plants can be estimated by comparing varieties released in different years grown contemporaneously from stored seed. Trials show that there was a steady increase of approximately 1 per cent in the yield of maize in the USA per year of introduction over a 70-year period since 1930 (Duvick *et al.* 2004).

Results for a limited number of generations are shown for cattle in figure 1, but the most intensive continuous selection in livestock has been practised in broiler chickens since the 1950s when specialist meat and egg lines were developed. Responses from selection based primarily on individual phenotype have been enormous (table 1), showing an approximately five-fold increase in 56-day body weight between 1957 and 2001 (Havenstein *et al.* 1994, 2003). Comparisons using modern and old diet formulations showed that at least 80 per cent of these differences were genetic. Responses were continuing at similar rates during the decade since 1991, other than in fatness, where selection to reduce fat had been effective (table 1). Intensive selection on specific traits has led to unfavourable changes in other characters, typically those that are associated with fitness, such as fertility in dairy cattle and leg strength and viability in poultry. Selection pressure has increasingly been put on such traits, such that in broilers viability and leg quality has improved in recent years (McKay *et al.* 2000; Havenstein *et al.* 2003; Hill & Zhang 2009). What these show is that the breeder has to be cognizant of all important traits; but if appropriate selection pressure can then be exerted, a change in direction can be effective, as the Illinois maize lines illustrate (figure 2).

It is not surprising that such continued responses are found, as in many other experiments (Hill & Bünger 2004). If many genes affect a trait, changes in gene frequency under selection are small, so variance is expected to change only slowly (Falconer & Mackay 1996); reductions in variance from those initially at high frequency may be largely compensated by increases in those initially rare; the influence of epistasis on response appears to be small (Crow 2008); and new models provide rationale (Barton & de Vladar 2009). Under the infinitesimal model, the total response deriving from the initial variation is expected to total $2N_e$ times the response in each

early generation (Robertson 1960). The new variation arising from continuous mutation, an increment in heritability of the order of 0.1 per cent per generation, implies that substantial continued responses can also be achieved from mutations. This has been demonstrated in selection experiments from inbred bases (Hill 1982; Keightley & Halligan 2009), although many mutations revealed in selection experiments are retained only because their effects on the selected trait outweigh those on viability or fertility (López & López-Fanjul 1993). Taken together, Walsh (2004) showed that the response in the Illinois lines was mainly contributed by variation in the founder lines, but must have been due partly to mutations arising subsequently.

Modern breeding programmes inevitably involve a concentration of improvement in populations of limited size so that effective multi-trait recording can be undertaken and intense selection practised. There is a multiplication pyramid from nucleus populations in poultry and pigs, and in dairy cattle a concentration through use of sires through artificial insemination worldwide. Breeding programmes can be designed to optimize the trade-off between high selection intensity with the use of relatives' information to increase short-term gain and the decrease in N_e and likely long-term progress (Villanueva *et al.* 2006). But are there problems?

For cattle there is evidence that population sizes were large following domestication, of the order of tens of thousands or more, but those in some modern breeds are of the order of 100. Even so the levels of molecular genetic diversity within breeds are at least as great as in human populations (Gibbs *et al.* 2009). Nucleus populations of chickens likely have similar effective sizes. An analysis by Muir *et al.* (2008) of a large collection of lines of broilers, layers and those maintained by fanciers indicated that about one-half of the alleles present in Red Jungle Fowl, regarded as the progenitor native population, had been lost, with most of it occurring in early years of domestication. Yet, heritability remains high, indeed that for body weight seems to maintain its traditional value of about 25 per cent regardless. Further, each year over 40×10^9 chickens are raised so, with a mutation rate of 1.8×10^{-9} , there are over 50 mutants at each DNA site. The problem is not that there is no new variation, but to identify the useful new variants.

Although it would be impossible to identify a mutant for a quantitative trait such as body weight in birds down the multiplication pyramid, it might be possible for a disease-resistant mutant.

We can be optimistic about the prospects for future improvement, not least because the input of molecular and high-throughput technologies to livestock improvement has so far been tiny. Clearly, there are limits imposed by the laws of thermodynamics, but by simply increasing the rate of live-weight gain of a bird, the efficiency of feed use is increased and also, a new consideration, greenhouse gas emissions per unit product is reduced. There are undoubtedly challenges, for example in the availability of water and climate change influences more generally, but new opportunities will come from new technology. Some, for example genomic selection, are really just extensions of classical quantitative genetic methods of increasing accuracy of selection. Others, for example changing or inserting new genes, provide radical ways of introducing new variation, but only if the public accepts them. Although conserved animal germ plasm far behind the commercial norm may harbour useful variants, I expect their contribution to genetic improvement to be small.

Similarly, the large amounts of genetic variation found in natural populations show that traits can be changed rapidly and substantially as a consequence of natural selection. With fitness defined as some simple measure, like bristle number in *Drosophila*, the effectiveness is illustrated by the results of many selection experiments (e.g. Weber 2004). There are also cases where fitness profiles and subsequently traits have changed greatly as a consequence of environmental change; for example, size of guppies increased substantially after transfer from a high to low predation environment, at rates similar to those found in laboratory selection experiments (Reznick *et al.* 1997). Although additive genetic variation and directional selection for particular traits have been shown, rarely have direct observations of natural populations revealed evolutionary changes, and those where responses were as expected were restricted to changes over one generation (Merila *et al.* 2001).

The ability to evolve depends on the additive genetic covariance structure of all the relevant traits, and whether the relevant combination actually expresses genetic variation. Recent analyses on genetic covariance matrices typically find that many of their eigenvalues are zero, such that the corresponding eigenvectors indicate directions of no variance (e.g. Blows & Walsh 2009; Kirkpatrick 2009; Walsh & Lynch 2009, ch. 30), which if these coincide with fitness ‘objectives’, implies adaptive evolution is not possible. Although these analyses indicate there are, indeed, trajectories that cannot be followed, sampling errors alone can lead to such inferences. To understand and predict changes or lack thereof, we greatly need more reliable information on the genetic covariances among multiple traits and on fitness profiles on many environments, but this is a massive task. In the presence of a major change of environment where fitness profiles change, the risk to a species seems more likely to come from other species-filling niches or

evolving more rapidly rather than from its total inability to adapt.

9. CONCLUDING REMARKS

Our level of understanding of many features of quantitative traits is quite rudimentary: what the genes do and how they interact, how their effects are distributed, the extent and magnitude of pleiotropic effects, the relations to overall fitness, and how and why is so much variation maintained? At this stage, however, we find that the many classical genes of small effect model explains many of the phenomena we observe and provides a basis for predictions of change. We can and are using the new information we get, however. But we should bear in mind that, as Darwin perceived, evolution succeeds through simple selection.

I am grateful to Loeske Kruuk, Ian McMillan, Peter Visscher, Anna Wolc, Naomi Wray and Xu-Sheng Zhang for helpful comments on drafts of the paper.

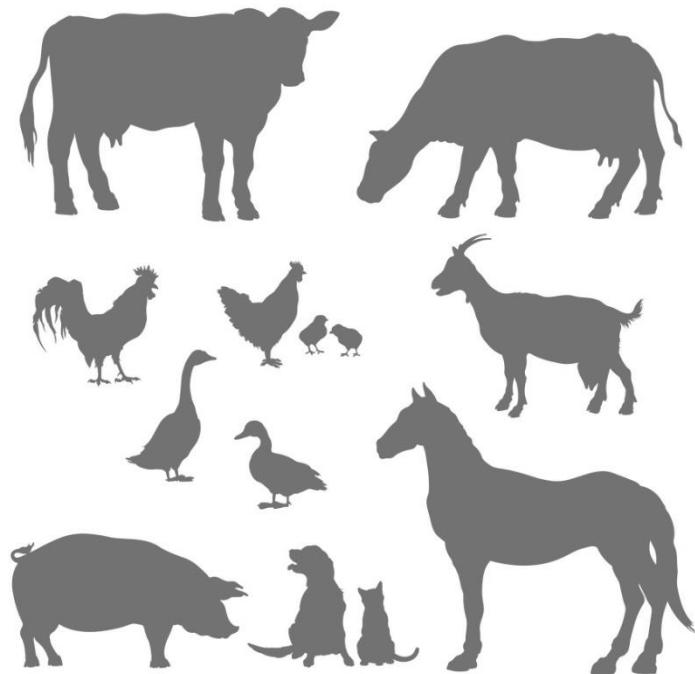
REFERENCES

- Baer, C. F. 2008 Quantifying the decanalizing effects of spontaneous mutations in rhabditid nematodes. *Am. Nat.* **172**, 272–281. (doi:10.1086/589455)
- Barton, N. H. & de Vladar, H. P. 2009 Statistical mechanics and the evolution of polygenic quantitative traits. *Genetics* **181**, 997–1011. (doi:10.1534/genetics.108.099309)
- Blows, B. & Walsh, B. 2009 Spherical cows grazing in flatland: constraints to selection and adaptation. In *Adaptation and fitness in animal populations, evolutionary and breeding perspectives on genetic resource management* (eds J. van der Werf, R. Frankham, H. Graser & C. Gondro), pp. 83–101. New York, NY: Springer.
- Bulmer, M. G. 1980 *The mathematical theory of quantitative genetics*. Oxford, UK: Oxford University Press.
- Bürger, R. 2000 *The mathematical theory of selection, recombination, and mutation*. New York, NY: Wiley.
- Carlberg, O. & Haley, C. S. 2004 Epistasis: too often neglected in complex trait studies? *Nat. Rev. Genet.* **5**, 618–625. (doi:10.1038/nrg1407)
- Chesler, E. J. *et al.* 2008 The collaborative cross at Oak Ridge National Laboratory: developing a powerful resource for systems genetics. *Mamm. Genome* **19**, 382–389. (doi:10.1007/s00335-008-9135-8)
- Clayton, G. A., Morris, J. A. & Robertson, A. 1957 An experimental check on quantitative genetic theory. 1. Short-term responses to selection. *J. Genet.* **55**, 131–151. (doi:10.1007/BF02981620)
- Cockerham, C. C. 1954 An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. *Genetics* **39**, 859–882.
- Collard, B. C. Y. & Mackill, D. J. 2008 Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Phil. Trans. R. Soc. B* **363**, 557–572. (doi:10.1098/rstb.2007.2170)
- Crow, J. F. 2008 Maintaining evolvability. *J. Genet.* **87**, 349–353. (doi:10.1007/s12041-008-0057-8)
- Dekkers, J. C. M. 2004 Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. *J. Anim. Sci.* **82**, E313–E328.
- Donnelly, P. 2008 Progress and challenges in genome-wide association studies in humans. *Nature* **456**, 728–731. (doi:10.1038/nature07631)

- Dudley, J. W. & Lambert, R. J. 2004 100 generations of selection for oil and protein content in corn. *Plant Breed. Rev.* **24**, 79–110.
- Duvick, D. N., Smith, J. S. C. & Cooper, M. 2004 Long-term selection in a commercial hybrid maize breeding program. *Plant Breed. Rev.* **24**, 109–151.
- Falconer, D. S. & Mackay, T. F. C. 1996 *Introduction to quantitative genetics*, 4th edn. Harlow, UK: Longman.
- Fisher, R. A. 1918 The correlation between relatives on the supposition of Mendelian inheritance. *Trans. R. Soc. Edin.* **52**, 399–433.
- Flint, J. & Mott, R. 2008 Applying mouse complex-trait resources to behavioural genetics. *Nature* **456**, 724–727. (doi:10.1038/nature07630)
- Fowler, K., Semple, C., Barton, N. H. & Partridge, L. 1997 Genetic variation for total fitness in *Drosophila melanogaster*. *Proc. R. Soc. Lond. B* **264**, 191–199.
- Frankham, R. 1990 Are responses to artificial selection for reproductive fitness characters consistently asymmetrical? *Genet. Res.* **56**, 35–42. (doi:10.1017/S0016672300028858)
- Gianola, D. & de los Campos, G. 2008 Inferring genetic values for quantitative traits non-parametrically. *Genet. Res.* **90**, 525–540. (doi:10.1017/S0016672308009890)
- Gibbs, R. A. et al. 2009 Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science* **324**, 528–532.
- Gimelfarb, A. & Willis, J. H. 1994 Linearity versus nonlinearity of offspring regression—an experimental study of *Drosophila melanogaster*. *Genetics* **138**, 343–352.
- Goddard, M. 2009 Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica* **136**, 245–257. (doi:10.1007/s10709-008-9308-0)
- Habier, D., Fernando, R. L. & Dekkers, J. C. M. 2009 Genomic selection using low-density panels. *Genetics* **182**, 343–353. (doi:10.1534/genetics.108.100289)
- Hadfield, J. D. 2008 Estimating evolutionary parameters when viability selection is operating. *Proc. R. Soc. B* **275**, 723–734. (doi:10.1098/rspb.2007.1013)
- Haley, C. S. & Knott, S. A. 1992 A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* **69**, 315–324.
- Havenstein, G. B., Ferket, P. R., Scheideler, S. E. & Larson, B. T. 1994 Growth, livability, and feed conversion of 1957 vs. 1991 broilers when fed ‘typical’ 1957 and 1991 broiler diets. *Poultry Sci.* **73**, 1785–1794.
- Havenstein, G. B., Ferket, P. R. & Qureshi, M. A. 2003 Growth, livability, and feed conversion of 1957 versus 2001 broilers when fed representative 1957 and 2001 broiler diets. *Poultry Sci.* **82**, 1500–1508.
- Hayes, B. J., Visscher, P. M. & Goddard, M. E. 2009 Increased accuracy of artificial selection by using the realised relationship matrix. *Genet. Res.* **91**, 47–60.
- Henderson, C. R. 1950 Estimation of genetic parameters. *Ann. Math. Stat.* **21**, 309–310.
- Henderson, C. R. 1984 *Applications of linear models in animal breeding*. Guelph, Ontario: University of Guelph.
- Hill, W. G. 1982 Predictions of response to artificial selection from new mutations. *Genet. Res.* **40**, 255–278. (doi:10.1017/S0016672300019145)
- Hill, W. G. & Bünger, L. 2004 Inferences on the genetics of quantitative traits from long-term selection in laboratory and farm animals. *Plant Breed. Rev.* **24**, 169–210.
- Hill, W. G. & Zhang, X. S. 2009 Maintaining genetic variation in fitness. In *Adaptation and fitness in animal populations, evolutionary and breeding perspectives on genetic resource management* (eds J. van der Werf, R. Frankham, H. Graser & C. Gondro), pp. 59–81. New York, NY: Springer.
- Hill, W. G., Goddard, M. E. & Visscher, P. M. 2008 Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genet.* **4**, e1000008. (doi:10.1371/journal.pgen.1000008).
- Hospital, F. 2008 Challenges for effective marker-assisted selection in plants. *Genetica* **136**, 303–310. (doi:10.1007/s10709-008-9307-1)
- Houle, D. 1992 Comparing evolvability and variability of quantitative traits. *Genetics* **130**, 195–204.
- Hu, X. X., Gao, Y., Feng, C. G., Liu, Q. Y., Wang, X. B., Du, Z., Wang, Q. S. & Li, N. 2009 Advanced technologies for genomic analysis in farm animals and its application for QTL mapping. *Genetica* **136**, 371–386. (doi:10.1007/s10709-008-9338-7)
- Johnson, T. & Barton, N. H. 2005 Theoretical models of selection and mutation on quantitative traits. *Phil. Trans. R. Soc. B* **360**, 1411–1425. (doi:10.1098/rstb.2005.1667)
- Keightley, P. D. & Halligan, D. L. 2009 Analysis and implications of mutational variation. *Genetica* **136**, 359–369. (doi:10.1007/s10709-008-9304-4)
- Keightley, P. D. & Hill, W. G. 1990 Variation maintained in quantitative traits with mutation-selection balance: pleiotropic side-effects on fitness traits. *Proc. R. Soc. Lond. B* **242**, 95–100. (doi:10.1098/rspb.1990.0110)
- Kempthorne, O. 1954 The correlation between relatives in a random mating population. *Proc. R. Soc. Lond. B* **143**, 102–113. (doi:10.1098/rspb.1954.0056)
- Kingsolver, J. G., Hoekstra, H. E., Hoekstra, J. M., Berrigan, D., Vignieri, S. N., Hill, C. E., Hoang, A., Gilbert, P. & Beerli, P. 2001 The strength of phenotypic selection in natural populations. *Am. Nat.* **157**, 245–261. (doi:10.1086/319193)
- Kirkpatrick, M. 2009 Patterns of quantitative genetic variation in multiple dimensions. *Genetica* **136**, 271–284. (doi:10.1007/s10709-008-9302-6)
- Kirkpatrick, M. & Heckman, N. 1989 A quantitative genetic model for growth, shape and other infinite-dimensional characters. *J. Math. Biol.* **27**, 429–450. (doi:10.1007/BF00290638)
- Kruuk, L. E. B. 2004 Estimating genetic parameters in wild populations using the ‘animal model’. *Phil. Trans. R. Soc. Lond. B* **359**, 873–890. (doi:10.1098/rstb.2003.1437)
- Kruuk, L. E. B., Clutton-Brock, T. H., Slate, J., Pemberton, J. M. et al. 2000 Heritability of fitness in a wild mammal population. *Proc. Natl Acad. Sci. USA* **97**, 698–703. (doi:10.1073/pnas.97.2.698)
- Lander, E. S. & Botstein, D. 1989 Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**, 185–199.
- Laurie, C. C. et al. 2004 The genetic architecture of response to long-term artificial selection for oil concentration in the maize kernel. *Genetics* **168**, 2141–2155. (doi:10.1534/genetics.104.029686)
- Legarra, A., Robert-Granié, C., Manfredi, E. & Elsen, M. 2008 *Genetics* **180**, 611–618. (doi:10.1534/genetics.108.088575)
- López, M. A. & López-Fanjul, C. 1993 Spontaneous mutation for a quantitative trait in *Drosophila melanogaster*. I. Distribution of mutant effects on the trait and on fitness. *Genet. Res.* **61**, 117–126. (doi:10.1017/S0016672300031220)
- Lynch, M. & Walsh, B. 1998 *Genetics and analysis of quantitative traits*. Sunderland, MA: Sinauer Associates.
- Mackay, T. F. C. 2009 The genetic architecture of complex behaviours: lessons from *Drosophila*. *Genetica* **136**, 295–302. (doi:10.1007/s10709-008-9310-6)
- Mackay, T. F. C. & Lyman, R. F. 2005 *Drosophila* bristles and the nature of quantitative genetic variation. *Phil. Trans. R. Soc. B* **360**, 1513–1527. (doi:10.1098/rstb.2005.1672)

- Mackay, T. F. C., Stone, E. A. & Ayroles, J. F. 2009 The genetics of quantitative traits: challenges and prospects. *Nat. Rev. Genet.* **10**, 565–577. (doi:10.1038/nrg2612)
- Maher, B. 2008 Personal genomes: the case of the missing heritability. *Nature* **456**, 18–21. (doi:10.1038/456018a)
- Martinez, V., Bünger, L. & Hill, W. G. 2000 Analysis of response to 20 generations of selection for body composition in mice: fit to infinitesimal assumptions. *Genet. Sel. Evol.* **32**, 3–21. (doi:10.1186/1297-9686-32-1-3)
- McKay, J. C., Barton, N. F., Koerhuis, A. N. M. & McAdam, J. 2000 The challenge of genetic change in the broiler chicken. In *The challenge of genetic change in animal production*, vol. 27 (eds W. G. Hill *et al.*), pp. 1–7. Edinburgh, UK: British Society of Animal Science.
- Merila, J., Sheldon, B. C. & Kruuk, L. E. B. 2001 Explaining stasis: microevolutionary studies in natural populations. *Genetica* **112**, 199–222.
- Meuwissen, T. H. E., Hayes, B. J. & Goddard, M. E. 2001 Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157**, 1819–1829.
- Meyer, K. & Hill, W. G. 1991 Mixed model analysis of a selection experiment for food intake in mice. *Genet. Res. Camb.* **57**, 71–81. (doi:10.1017/S0016672300029062)
- Mousseau, T. A. & Roff, D. 1987 Natural selection and the heritability of fitness components. *Heredity* **59**, 181–197. (doi:10.1038/hdy.1987.113)
- Muir, W. M. *et al.* 2008 Genome-wide assessment of worldwide chicken SNP genetic diversity indicates significant absence of rare alleles in commercial breeds. *Proc. Natl Acad. Sci. USA* **105**, 17312–17317. (doi:10.1073/pnas.0806569105)
- Pearson, H. 2009 Human genetics: one gene, twenty years. *Nature* **460**, 164–169. (doi:10.1038/460164a)
- Purcell, S. M. *et al.* 2009 Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–752.
- Reznick, D. N., Shaw, F. H., Rodd, F. H. & Shaw, R. G. 1997 Evaluation of the rate of evolution in natural populations of guppies (*Poecilia reticulata*). *Science* **275**, 1934–1937. (doi:10.1126/science.275.5308.1934)
- Robertson, A. 1960 A theory of limits in artificial selection. *Proc. R. Soc. Lond. B* **153**, 234–249. (doi:10.1098/rspb.1960.0099)
- Robertson, A. 1967 The nature of quantitative genetic variation. In *Heritage from Mendel* (ed. A. Brink), pp. 265–280. Madison, WI: University of Wisconsin Press.
- Roff, D. A. 2007 A centennial celebration for quantitative genetics. *Evolution* **61**, 1017–1032. (doi:10.1111/j.1558-5646.2007.00100.x)
- Schaeffer, L. R. & Dekkers, J. C. M. 1994 Random regressions in animal models for test-day production in dairy cattle. In *Proc. 5th World Congr. Genet. Appl. Livest. Prod.*, vol. 18, pp. 443–446. Ontario, Canada: Guelph.
- Sheridan, A. K. 1988 Agreement between estimated and realised genetic parameters. *Anim. Breed. Abstr.* **56**, 877–889.
- Slatkin, M. 2009 Epigenetic inheritance and the missing heritability problem. *Genetics* **182**, 845–850. (doi:10.1534/genetics.109.102798)
- Sorensen, D. & Gianola, D. 2002 *Likelihood, Bayesian and MCMC methods in quantitative genetics*. New York, NY: Springer.
- Sorensen, D. & Waagepetersen, R. 2003 Normal linear models with genetically structured residual variance heterogeneity: a case study. *Genet. Res. Camb.* **82**, 207–222. (doi:10.1017/S0016672303006426)
- Thompson, R. 2008 Estimation of quantitative genetic parameters. *Proc. R. Soc. B* **275**, 679–686. (doi:10.1098/rspb.2007.1417)
- Turelli, M. 1984 Heritable genetic variation via mutation-selection balance: Lerch's zeta meets the abdominal bristle. *Theoret. Popul. Biol.* **25**, 138–193. (doi:10.1016/0040-5809(84)90017-0)
- Valdar, W. *et al.* 2006 Genome-wide genetic association of complex traits in heterogeneous stock mice. *Nat. Genet.* **38**, 879–887. (doi:10.1038/ng1840)
- Van Raden, P. M., Van Tassell, C. P., Wiggans, G. R., Sonstegard, T. S., Schnabel, R. D., Taylor, J. F. & Schenkel, F. S. 2009 Reliability of genomic predictions for North American Holstein bulls. *J. Dairy Sci.* **92**, 16–24. (doi:10.3168/jds.2008-1514)
- Villanueva, B., Avendaño, S. & Woolliams, J. A. 2006 Prediction of genetic gain from quadratic optimization with constrained rates of inbreeding. *Genet. Sel. Evol.* **38**, 127–146.
- Visscher, P. M. 2008 Sizing up human height variation. *Nat. Genet.* **40**, 489–490. (doi:10.1038/ng0508-489)
- Visscher, P. M., Medland, S. E., Ferreira, M. A. R., Morley, K. I., Zhu, G., Cornes, B. K., Montgomery, G. W. & Martin, N. G. 2006 Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. *PLoS Genet.* **2**, e41. (doi:10.1371/journal.pgen.0020041)
- Visscher, P. M. *et al.* 2007 Genome partitioning of genetic variation for height from 11,214 sibling pairs. *Am. J. Hum. Genet.* **81**, 1104–1110. (doi:10.1086/522934)
- Wagner, G. P., Kenney-Hunt, J. P., Pavlicev, M., Peck, J. R., Waxman, D. & Cheverud, J. M. 2008 Pleiotropic scaling of effects and the ‘cost of complexity’. *Nature* **452**, 470–473. (doi:10.1038/nature06756)
- Walsh, B. 2004 Population- and quantitative-genetic analyses of selection limits. *Plant Breed. Rev.* **24**, 177–225.
- Walsh, B. & Lynch, M. 2009 Genetics and analysis of quantitative traits. Volume 2: Evolution and selection of quantitative traits. See http://nitro.biosci.arizona.edu/zbook/NewVolume_2/newvol2.html.
- Weber, K. 2004 Population size and long-term selection. *Plant Breed. Rev.* **24**, 249–268.
- Weedon, M. N. & Frayling, T. M. 2008 Reaching new heights: insights into the genetics of human stature. *Trends Genet.* **24**, 595–603. (doi:10.1016/j.tig.2008.09.006)
- Weller, J. I. 2009 *Quantitative trait loci analysis in animals*, 2nd edn, p. 272. Wallingford, Oxford: CABI.
- Wray, N. R., Goddard, M. E. & Visscher, P. M. 2007 Prediction of individual genetic risk to disease from genome-wide association studies. *Genome Res.* **17**, 1520–1528. (doi:10.1101/gr.6665407)
- Wright, S. 1921 Systems of mating. Parts I–V. *Genetics* **6**, 111–178.
- Wright, S. 1931 Evolution in Mendelian populations. *Genetics* **16**, 97–159.
- Zhang, X.-S. & Hill, W. G. 2005a Predictions of patterns of response to artificial selection in lines derived from natural populations. *Genetics* **169**, 411–425. (doi:10.1534/genetics.104.032573)
- Zhang, X.-S. & Hill, W. G. 2005b Genetic variability under mutation selection balance. *Trends Ecol. Evol.* **20**, 468–470.
- Zhang, X.-S. & Hill, W. G. 2008 Mutation selection balance for environmental variance. *Am. Nat.* **171**, 394–399. (doi:10.1086/527503)

3. LIVESTOCK



3.1 Study book



Genetic Improvement of Farmed Animals

Geoff Simm, Geoff Pollott, Raphael Mrode,
Ross Houston and Karen Marshall



Title: ***Genetic Improvement
of Farmed Animals***

Edition: 1

Authors: **Geoff Simm,
Geoff Pollott,
Ross Houston,
Raphael A Mrode, and
Karen Marshall**

Year: 2020

Publisher: **Cabi Publishing**

3.1.1

Simm et al. (2020)

Page 43-63

3. Strategies for Genetic Improvement

Introduction

For thousands of years humans have attempted to alter populations of animals to make them more suitable for the production of food and fibre, as providers of transport, draught power, etc. These attempts have been increasingly effective over the last couple of centuries, especially since the mid-1900s. Generally, improved breeds of livestock produce food, fibre or other products, which are of higher quality, or are better matched to modern requirements than their predecessors. Also, improved breeds usually have higher efficiency of production than unimproved breeds, and so the relative cost of production is lower. For example, in many countries genetic improvement of pigs and poultry has contributed to the change in status of pig and poultry meat from being luxury foods, to being the cheapest meats available. In many countries, the price of milk has fallen in real terms over the last few decades, partly as a result of genetic improvement of yield and overall efficiency of production. Selection for reduced fibre diameter and increased fleece weight in specialized sheep breeds has been important in allowing wool to continue to win a share of the market for clothing fabrics. Genetic improvement is particularly valuable because it is permanent, it is cumulative when selection is continuous, and it is usually highly cost effective (we give examples in later chapters on different farmed species).

The aim of this chapter is to describe the strategies which are employed to achieve genetic improvement. Some more detailed examples of both the benefits and, importantly, the possible negative consequences of selection, and how these can be avoided or reduced, are discussed in later chapters. Traditionally three main strategies have been used for the genetic improvement of livestock. These are:

- *Selection between breeds* or breed types – comparing two or more breed types and selecting the most appropriate, or substituting one breed type for another.
- *Selection within breeds* or breed types – choosing better parents within a particular breed type.
- *Crossbreeding* – mating parents of two or more different breed types or species together.

We use the terms breed type here to cover conventional pure breeds, strains within breeds, composite or synthetic lines, and populations recently derived from wild stock, as in the case of finfish.

These strategies are discussed in more detail below. Developments in molecular and reproductive biology are augmenting these approaches, e.g. the growing use of markers in *genomic selection* within breeds or strains. These developments are also allowing new strategies, still largely in experimental use in livestock, including the ability to transfer genes within or between species, to 'edit' genes, and to regulate or modify the expression of existing or introduced genes. These new possibilities are discussed in more detail in Chapter 5.

For any genetic improvement strategy to be effective, it is important to have a clear view of what the economically – or otherwise – important animal characteristics (traits) are. Then it is logical to choose the most appropriate breed type or cross, based on their performance in these traits. It is then sensible to consider whether this pure breed, or component breeds of the chosen cross, can be improved further by *within-breed selection*. In practice, the availability of information on the performance of different breeds and crosses, the financial and physical resources available, the current breeds or crosses in use, local traditions, local market demands and personal preferences also influence the choice of breed or cross, and the extent to which within-breed selection is practised, at least in the short term.

Before discussing the different strategies for genetic improvement in detail, it will be helpful to consider the structure of the livestock breeding industries in which they are applied.

The Structure of Livestock Breeding Industries

The structure of livestock breeding industries in most industrialized nations is often described schematically as a pyramid with *elite* (or *nucleus*, *seedstock* or *stud*) populations or breeders at the top, one or more middle tiers of purebred or *crossbred* multipliers, and a final tier of commercial herds or flocks, or end users (see Fig. 3.1). In this generalized model, the elite breeders' role is to produce breeding stock (particularly males, as fewer of these are needed to produce a given number of offspring) for use within the top tier and in multipliers' herds or flocks. The elite populations are the main focus for genetic improvement. The main rôle of multipliers, as the name suggests, is to take improved stock from the tier above, and to create larger numbers of breeding animals for sale to the tier below. In most industries, there are purebred multipliers involved in producing greater numbers of purebred animals, particularly males, for the tiers below. However, in some industries there are also crossbred multipliers, who produce crossbred animals, particularly females, for use in the commercial tier. Usually flocks or herds in the commercial tier are primarily involved in the production of eggs, meat, milk or fibre, and have little or no involvement in selling stock for further breeding. Hence, genetic improvement flows down the pyramid, or is disseminated, from elite, through multiplier, to commercial tiers. The difference in genetic merit of different tiers of the pyramid is termed the *improvement lag*.

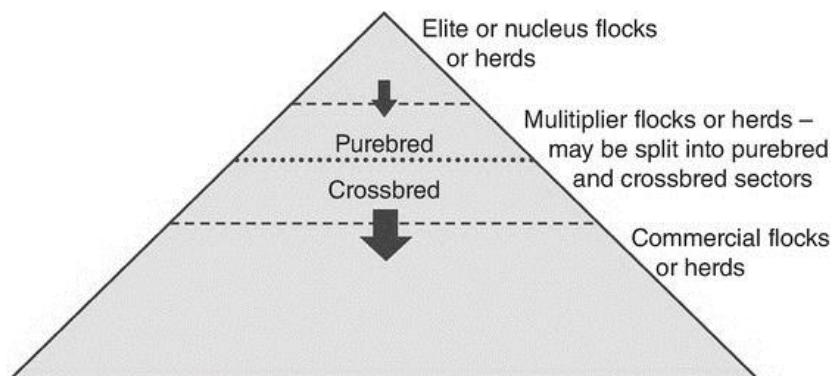


Fig. 3.1. Diagram illustrating the structure of livestock breeding industries in many industrialized nations. This is a pyramid with elite or nucleus populations at the top, one or more middle tiers of purebred or crossbred multipliers, and a final tier of commercial herds or flocks, or end users. (After Nicholas, 1987.)

In the pig and poultry industries in many countries, the elite and multiplier tiers are mainly in the hands of a small number of multinational breeding companies. They maintain elite purebred or synthetic lines, often in several countries, in which most of the selection is practised. Breeding stock from these lines then moves to the company's own multiplier herds or flocks, or to herds or flocks contracted by them, to produce both males and females for sale to commercial producers. These pigs and poultry produced for the commercial tier are usually crossbred. (In some countries there is further 'vertical integration', with companies being involved from nucleus breeding, multiplication through to commercial production and processing.)

For the most advanced aquaculture sectors, such as Atlantic salmon, the supply of genetically improved stock is largely controlled by few, large multinational breeding companies (in some cases, the same companies that produce elite pigs and poultry lines). There is less need for multiplication tiers for aquaculture species, and large breeding nuclei can supply siblings of elite *broodstock* for production, or they have a single multiplication layer. The companies do not typically produce specialized lines, nor is crossbreeding common, but some companies do offer *marker-assisted* or genomic selection for specific characteristics (e.g. resistance to specific pathogens). For the aquaculture sectors that are less advanced, the status of breeding programmes and the commercial interest varies dramatically within a single species. For example, Nile tilapia is arguably the world's most important food fish, and while some genetically improved lines are bred and disseminated globally by major multinational breeding companies, the majority of production is via local hatcheries and smallholder farmers. In these systems the breeding is less well controlled and can be a mix of commercial and public sector suppliers of stock. Across most sectors, there is a mix of specialized breeding companies with core business in supply of genetically improved eggs and juveniles and also integrated companies who perform both selective breeding and production via vertical integration, as observed in some pig and poultry companies.

The very widespread use of artificial insemination (AI) in the dairy industries of most countries has effectively removed the middle tier from the breeding pyramid. That is, the commercial tier has direct access to elite animals via AI. The elite tier in dairy cattle breeding is usually owned by a mixture of private breeders and breeding companies – the former own most of the elite cows, and the latter own most of the elite bulls. Unlike the situation in the pig and poultry industries, in most temperate countries the commercial tier in the dairy industry is made up largely of purebred animals (New Zealand is a notable exception).

In most countries, all tiers of the pyramid in the sheep, goat and beef cattle breeding industries are in the hands of individual breeders, rather than breeding companies. (A few sheep and beef breeding companies have emerged in the last few years and have gained a significant market share in New Zealand, North and South America.) However, the structures of the breeding industries vary quite markedly from country to country. For instance, the Australian finewool industry is based on purebred Merino sheep in all tiers. The middle and lower tiers of the Merino industry also act as crossbred multipliers for some sectors of the lamb meat industry, by producing half-bred Merinos (crosses between Merino ewes and, usually, Border Leicester rams). These, in turn, are mated to rams from *terminal sire* or meat breeds (especially the Poll Dorset). The New Zealand sheep industry was based largely on purebred dual-purpose ewes in all tiers. These included the Romney, Coopworth, Perendale and Corriedale breeds which had been selected to produce both meat and wool. In commercial flocks a proportion of these pure dual-purpose ewes were often crossed to terminal sires. With stronger market demand for meat than wool in the last few decades, there has

been a shift in both Australia and New Zealand towards more specialized meat-producing breeds or crosses, including newly created crosses or composites. In the UK, hill sheep flocks in all tiers are usually purebred. As in the Merino example, many of these UK hill flocks act as crossbred multipliers, producing crossbred ewes for the commercial tier of the sheep industry in the lowlands and uplands. These are usually produced from *draft* hill ewes – those which have had about four lamb crops on the hill and have been moved to a less extreme environment for crossing, usually to rams from one of the longwool breeds. The resulting crossbred ewes are generally mated to purebred rams from a terminal sire or specialized meat breed, to produce lambs for meat production.

The pastoral beef industries of many temperate countries or regions are based on purebred cows of the traditional British beef breeds – Hereford, Aberdeen Angus and Shorthorn, or crosses between them. Often these are mated to purebred bulls from the larger continental European breeds, such as the Charolais, Simmental or Limousin. Purebred animals are the norm in all tiers of the French beef industry. In contrast, much of the commercial tier in Britain and Ireland has been based traditionally on beef × dairy crossbred cows. Until recently, the popularity of these crossbred cows was partly because of the plentiful supply of replacement beef × dairy heifers from dairy herds – a by-product of the practice of crossing dairy heifers, and those cows not required for breeding replacement dairy heifers, to a beef bull. In these cases, dairy herds act as multipliers of crossbred cows for the beef industry. (These cows kept for rearing calves for beef production are called *suckler* cows.) Table 3.1 gives some examples of where some typical livestock enterprises fit on the breeding and dissemination pyramid.

The terms purebred and pedigree require further explanation. Purebred or straightbred animals are those produced from several or many generations of matings between animals of the same breed. Strictly speaking, any purebred animal whose parentage, or more distant ancestry, is known can be described as a *pedigree animal*. However, pedigree is often used to mean *pedigree registered* – in other words an animal whose ancestry, and other details, are officially recorded by a *breed society* or other organization set up for this purpose. In Britain, most sheep in the elite tier, except in some hill breeds, and virtually all beef and dairy cattle in the elite tier are pedigree registered. This is also true for a significant proportion of ‘commercial’ dairy cattle. Most purebred beef cattle and sheep in the multiplier tier are pedigree registered; others have unofficial records of ancestry but are not registered with a breed society. In the commercial tier most beef and sheep are crossbred, though a few are purebred but not pedigree registered. Traditionally, great status has been attached to animals which are officially pedigree registered. In many cases pedigree-registered animals have a higher monetary value than non-registered animals, though this does not always reflect their genetic merit. Breed societies have generally provided a valuable service in maintaining accurate records of animals’ pedigrees – a service which is valued by both breeders and potential buyers of pedigree livestock. However, until relatively recently, fewer societies have actively promoted performance recording, and the use of both pedigree and performance records to assist selection decisions. The combination of performance and pedigree records is one of the factors which makes modern methods of genetic improvement so effective. Increasingly, breed societies have recognized this, and many have broadened their services to provide detailed information on performance as well as ancestry, and have improved their members’ access to, and understanding of, selection tools.

Table 3.1. Examples of enterprises in various tiers of the 'breeding pyramid'. Several tiers may be present in a single farm or business.
(After Simm et al., 1994 and Simm, 1998.)

Tier	Dairy cattle	Beef cattle	Sheep	Pigs	Poultry
1. Elite or nucleus breeders	Breeding companies 'contract mating' top AI bulls to elite cows often in private ownership. Pedigree Holstein Friesian breeders selling heifers to other breeders in this tier, or to the tier below.	Pedigree terminal sire (or other) beef breeders, selling bulls to other elite pedigree herds (tier 1), and purebred multiplier herds (tier 2).	Pedigree terminal sire or longwool sheep breeders, selling rams to other elite pedigree flocks (tier 1) and multiplier pedigree flocks (tier 2). Purebred 'stud' breeders of Merinos selling rams within this tier and to purebred multipliers.	Breeding companies selecting sire and dam lines. Pedigree breeders.	Breeding companies selecting egg laying and broiler strains of chicken. Pedigree breeders.
2. Purebred multipliers	Pedigree or other breeders producing heifers for sale to commercial herds.	Pedigree terminal sire herds buying bulls from tier 1, and selling bulls for crossing in commercial herds (tier 4).	Pedigree terminal sire flocks buying rams from tier 1, and selling rams for crossing in commercial flocks (tier 4). Pedigree longwool breeders selling rams to crossbred multipliers.	Breeding companies multiplying sire and dam lines for sale to tier 4. Pedigree breeders.	Breeding companies multiplying egg laying and broiler strains. Pedigree breeders.
3. Crossbred multipliers	-	Dairy herds buying beef bulls from tier 2 or beef semen from tier 1 or 2, and selling beef x dairy heifers to suckler herds in tier 4. Pure beef herds (e.g. Galloway) crossing to another beef breed (e.g. White Shorthorn) and selling crossbred heifers (e.g. Blue Grey) to suckler herds in tier 4.	Draft ('retired') hill ewes (e.g. Scottish Blackface) crossed to longwool sires (e.g. Bluefaced Leicester) to produce crossbred breeding females (e.g. Scottish Mule) for use in tier 4. Draft Merino ewes crossed to Border Leicester rams to produce crossbred females for tier 4.	Breeding companies or their partners/agents creating crosses for sale to tier 4 producers.	Breeding companies or their partners/agents creating crosses for sale to tier 4 producers.
4. Commercial or end users	Purebred Holstein Friesian (or other) dairy herds, using AI with semen from bulls in tier 1.	Crossbred suckler cow herds (e.g. Hereford or Simmental x Holstein Friesian; Blue Grey) buying replacement females from tier 3, and bulls from tier 2.	Crossbred ewe flock (e.g. Scottish Mule, Border Leicester x Merino) buying replacement females from tier 3, and rams from tier 2.	Commercial pig farmers buying replacement gilts and boars from multipliers in tier 3.	Commercial egg producers buying point-of-lay pullets from multipliers; commercial broiler producers buying day old chicks from multipliers/hatcheries in tier 3.

In the idealized model of a livestock breeding industry, strong market signals ought to go up this pyramid from one tier to the next, resulting in a clearly defined *breeding goal* in each tier. This is an accurate reflection of what happens in most pig, poultry, finfish and dairy cattle breeding industries in industrialized countries. It is also increasingly the case in beef cattle and sheep breeding industries, though in some breeding goals are still set ‘top down’ instead of ‘bottom up’. In other words, breeding goals for elite animals are influenced much more by market signals from within the top tier (e.g. by the high prices other breeders pay for show-winning or ‘fashionable’ animals), rather than from tiers below. Frustration with this situation led to the formation of *group breeding schemes* in several sheep and beef cattle breeds in New Zealand in the 1960s (Parker and Rae, 1982). The farmers concerned were not satisfied that the tiers above were producing breeding stock relevant to their needs, and so they established co-operative breeding schemes to produce their own replacement breeding stock, especially males. In most cases, this involved the formation of nucleus flocks or herds. These are elite flocks or herds formed by screening the best cows or ewes from co-operators’ own herds or flocks. Comprehensive recording and rigorous selection in the nucleus results in genetic improvement of commercially important traits in the nucleus animals, which is then disseminated to members’ own flocks or herds when they obtain replacement rams and bulls (or females) from the nucleus (see Fig. 3.2). While these schemes have become less common in industrialized countries, they are being applied to good effect in some lower-income countries, circumventing the lack of industry-wide infrastructure for breeding programmes. They often form the basis of community-based breeding programmes in these countries (see e.g. Mueller *et al.*, 2015 and later chapters).

A number of factors are involved in diluting the signals between tiers in livestock industries. The many links in the chain connecting primary producers to consumers often leads to imperfect communication, and so the signals reaching commercial producers may be weak in the first place. In some countries, direct or indirect subsidies may mask ‘true’ market signals. There may be a lack of understanding of, or low availability of, objective information on the genetic merit of breeds, crosses or individual animals within breeds. Also, social factors are often important – for instance, especially historically, some breeders preferred to breed show winners than to follow strict market signals. For a small and decreasing number of elite breeders there are still niche markets which are based mainly on show performance. At least in the short term, it is probably more profitable for this minority of breeders to breed for show performance, than to breed for the traits of most economic or other importance in the tiers below. Unfortunately, other breeders may be lured by the potentially big rewards from these small niche markets and neglect the larger markets. A further complication occurs if the animals produced in some tiers are by-products of other primary activities. For example, the main concerns of most dairy farmers crossing to a beef bull is the bull’s effect on calving ease in their dairy herd, and then the potential sale value of the calf. They are less concerned about the possible use of some daughters of the beef bull as suckler cows, and so most of them pay little attention to the performance traits which might be useful in that rôle (e.g. intermediate mature size and good maternal performance) when choosing a beef bull.

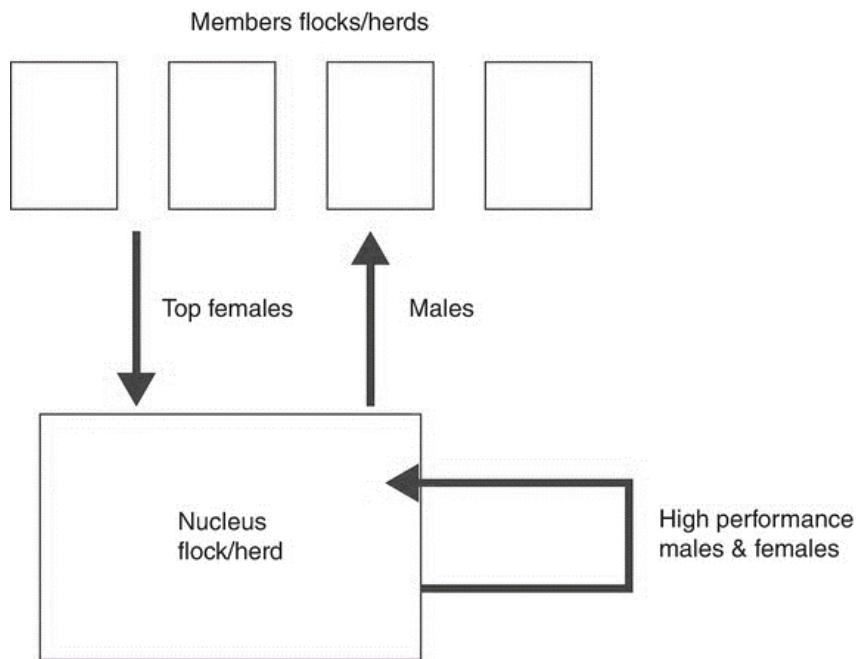


Fig. 3.2. A diagram illustrating the principles of group breeding schemes. These include (i) formation of a central nucleus herd or flock, from the best females available in members' own herds or flocks; (ii) comprehensive recording and rigorous selection for commercially-important traits in the nucleus; (iii) the best males produced in the nucleus are retained there for breeding, while the next best group (or ex-nucleus males) are the main source of breeding males for members' herds or flocks – this is the main route for dissemination of genetic improvement from the nucleus to the base herds or flocks; (iv) elite females may continue to be 'promoted' to the nucleus from base herds or flocks, on the basis of genetic merit (this only takes place in 'open' nucleus schemes – in some cases 'closed' nucleus schemes are preferred, for example to minimize risks of transfer of disease to the nucleus, or because of a lack of recording in base herds or flocks).

It will be apparent now that the number of tiers present in the breeding pyramid, and their relative importance, will vary between species and countries. The most relevant strategy for genetic improvement for any individual breeder will depend on the breeder's position in the breeding pyramid. Most elite breeders and purebred multipliers will be interested in within-breed selection, and occasionally in *between-breed selection*, if the market changes substantially and their current breed becomes less relevant. Crossbred multipliers will be interested mainly in crossbreeding. Commercial producers will probably be interested in several or all of the strategies for genetic improvement. Although they are users, rather than creators, of genetically improved stock, a knowledge of the techniques involved should allow them to make better decisions on which breed, cross or individual animals to use.

Selection Between Breeds or Breed Types

Selection between breeds or breed types can achieve dramatic and rapid genetic change when there are large genetic differences between populations in characteristics of economic and other importance. *Breed substitution* has led to major step changes in genetic improvement in numerous livestock production systems. However, this benefit is achieved only once per breed substitution, unlike the improvement brought about by continuous within-breed selection.

It is obviously costly to replace whole flocks or herds of breeding females at once. In practice, changes are often made more gradually. If replacement females (i.e. females brought into the breeding herd or flock to replace those that have died of natural causes, or have been *culled*) are usually purchased, then the switch to the new breed type can be made gradually by buying replacements from the new breed. If replacement females are normally homebred, then the composition of the herd or flock can be changed gradually by *grading up* or repeated *backcrossing* to the new breed. Grading up or backcrossing involves repeated mating of the current females, and subsequently their female offspring, to sires of the new breed. Table 3.2 shows the proportion of genes coming from each of two breeds of cattle or sheep, following successive generations of backcrossing – on average each generation of crossing halves the proportion of genes from the original breed, compared to that in the previous generation. Table 3.2 also shows the minimum time required to achieve particular proportions of genes from each breed. Many pedigree breed societies allow registration of crossbred animals in a separate register, with eventual acceptance as purebred animals once they are 7/8 (the EU legal standard) or 15/16 purebred. Modern pig, poultry and fish (and increasingly cattle and sheep) breeding operations are not subject to these strictures or traditions.

The first important criterion for choosing between breeds, strains or crosses is that the choice ought to be made on the basis of objective comparisons of performance *in the relevant environment* and at a relevant endpoint. There are some dramatic examples of the cost of ignoring this rule – high-performing temperate breeds of livestock have been introduced often to the tropics

Table 3.2. The proportion of genes from each of two breeds following successive generations of backcrossing, and the time taken to achieve these proportions in cattle and sheep. The letters A and B are used as abbreviations for the two breeds here (not as abbreviations for alleles, as in the previous chapter). In the first cross exactly half of each animal's genes come from each of the two breeds. In subsequent generations of backcrossing the proportions of genes shown are the averages expected – individual animals may have greater or smaller proportions than shown because of segregation and recombination.

Breed composition of:			Number of years after initial cross to produce offspring of the composition shown, assuming females first leave offspring at:		
Sire	Dam	Offspring	1 year old (some sheep breeds)	2 years old (sheep and some cattle breeds)	3 years old (cattle)
A	B	½ A, ½ B	–	–	–
A	½ A, ½ B	¾ A, ¼ B	1	2	3
A	¾ A, ¼ B	½ A, ½ B	2	4	6
A	½ A, ½ B	¹⁵/₁₆ A, ¹/₁₆ B	3	6	9

without this sort of trial, and have then succumbed to diseases or to nutritional deprivation, to which local breeds were tolerant. See Marshall (2014) and the International Livestock Research Institute

(ILRI) and partners' African Chicken Genetic Gains (ACGG, 2019) and African Dairy Genetic Gains (ADGG, 2019) projects, for more on this.

Less dramatically, there are important economic benefits to be gained, in any country, from matching breeds or crosses to particular production systems – as indicated by the old adage ‘horses for courses’ (see Fig. 3.3). The concept that genotypes do not always rank the same in different environments, or that the advantage to a particular genotype in one environment may be smaller or greater in another, is an important one in livestock improvement. In general, this is called a *genotype × environment (GxE) interaction*, or in this particular case, a *breed × production system interaction*. Figure 3.4 illustrates two types of genotype × environment interaction. The genetic correlation between the performance of related animals in two environments can also be used to detect genotype × environment interactions. In theory, if the genetic correlation between performance in the two environments is less than 1, then there is an interaction. However, correlations have to be substantially less than 1 before an interaction is of practical importance.

In wealthier countries, a great deal of research has been done on comparing the performance of different genotypes (breeds, strains, crosses, or animals of different genetic merit within a breed) in different environments (countries, production systems, feeding levels) to help livestock producers

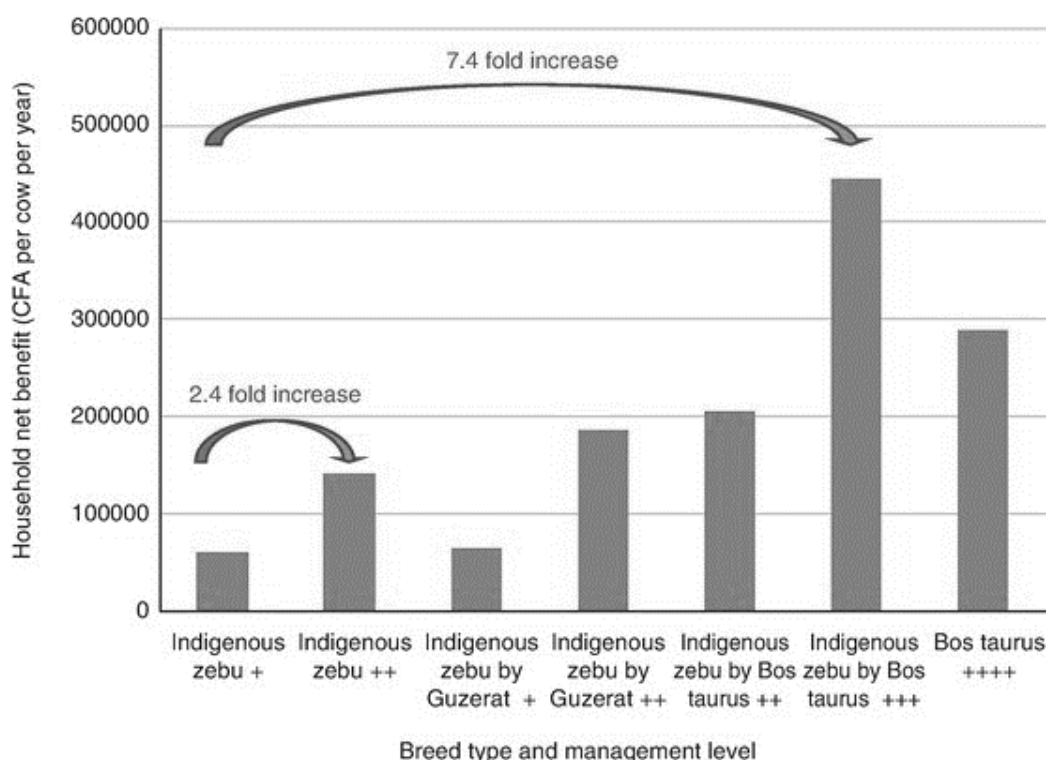


Fig. 3.3. Results of a survey on the impact of choice of dairy breed type and management level on household income in Senegal. The graph shows net benefit in West African CFA Franc per cow per annum, assuming a herd size of eight cows, with management level ranging from + (the lowest level of management) to +++++ (the highest level of management). (After Marshall et al., 2016, 2020.)

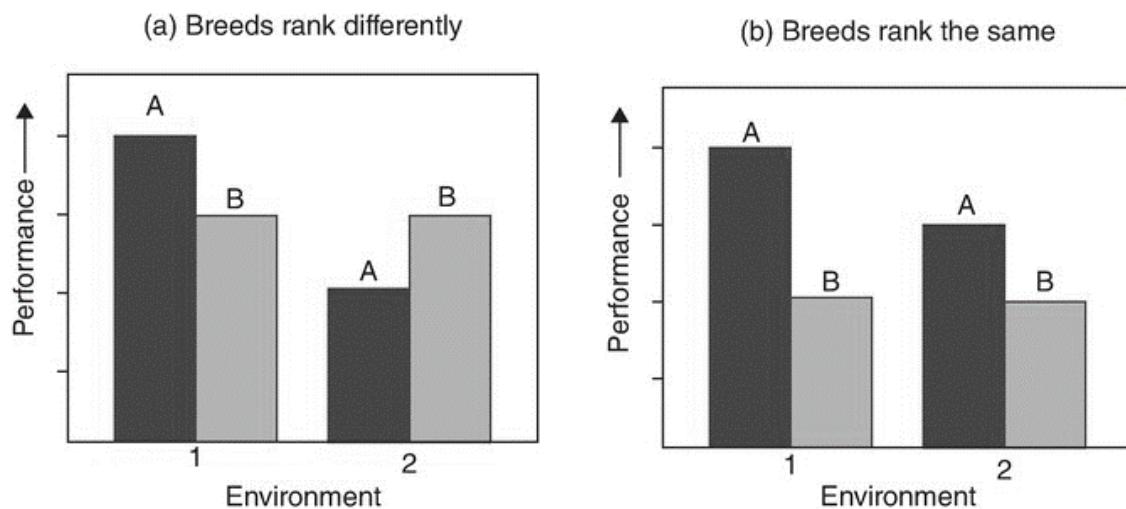


Fig. 3.4. A diagram illustrating two types of genotype \times environment interaction. (a) Breeds A and B rank differently in the two environments. (b) Breeds A and B rank the same in the two environments, but the advantage to breed A is much less in environment 2 than in environment 1.

choose the most appropriate breed type for their circumstances. In lower-income countries, however, much work remains to be done (Marshall, 2014). At first sight the results of many GxE studies are in conflict – many of them appear to deny the existence of interactions, while others firmly support their existence. It is probably fair to say that if there is a big enough difference in performance between the genotypes, or a big enough difference between the environments in which they are compared, or both, then there will be an interaction. For example, several classical studies indicate that small- or medium-sized beef and sheep breeds have higher overall productivity or profitability in extensive grazing systems than larger breeds. But larger breeds tend to do best in more intensive systems with high levels of concentrate feeding (Dickerson, 1978; Cartwright, 1982). There are several possible explanations for this. The more extensive the production system, the greater the effort animals will have to make to harvest the resources they require for maintenance and growth. Hence, smaller breeds may have an advantage in extensive systems, because they have lower requirements for maintenance and growth than larger breeds, and so their requirements are more in balance with the effort they expend in meeting them. An alternative explanation is that the larger sheep and beef breeds are probably the ones which have been subjected to the greatest artificial selection. At the same time, there may have been modifications to their environment, which have reduced natural selection for adaptation to extensive conditions. As a consequence of these two factors, favourable attributes for extensive systems (e.g. high mobility, particular patterns of grazing behaviour, cold or heat tolerance, drought tolerance, disease resistance and strong maternal behaviour) may have been lost, or may have diminished in larger, more intensely selected sheep and beef cattle breeds. (The results of some experiments investigating interactions are given in later chapters.)

Another important criterion for breed or breed type comparisons is that the samples of animals compared should be sufficiently large and representative of the breeds concerned to allow inferences to be drawn about these breeds in general, and not just the sample itself. In other words, the breed comparisons need plenty of animals from different sires and dams (though dams are usually automatically well represented in the less prolific species), and these sires themselves need to be sampled randomly, or in a balanced way, from several to numerous different flocks or herds. This

criterion is often violated – newly introduced breeds are often promoted on the basis of very flimsy comparisons, or no objective comparisons at all. The size of the sample required depends on the minimum difference between breeds that is considered to be of economic importance. It also depends on the amount of variation that exists in the trait concerned.

For example, let us assume that a 25 kg advantage in the 18-month weight of Charolais-cross beef calves was the minimum that would justify a producer switching from Simmental crosses. If the standard deviation of 18-month weight is about 50 kg in both crosses, then we would need to compare about 70 animals of each cross to be 90% sure that Charolais crosses were really better, and about 85 animals of each cross to be 95% sure. (See Snedecor and Cochran (1989) or other standard statistics textbooks, for details of how to calculate the numbers of animals required for this sort of test.)

Generally, the higher the degree of genetic control of a trait (i.e. the higher the heritability – this is the ratio of additive genetic to total phenotypic variation in the trait of interest), the greater the number of sires that should be used in a breed comparison. As a rough guide, five or more sires should be used if the traits of interest are lowly heritable, and ten or more should be used if the trait is highly heritable. (This seems odd at first sight, but there is proportionally more variation between sire families for highly heritable traits than for lowly heritable traits. This means that there is a greater risk of mistakenly choosing one breed over another, as a result of the chance sampling of one or two exceptional sires, when the heritability of the trait is high.) Ideally, sires should be from different flocks or herds. References to examples of well-designed breed comparisons are given at the end of this chapter. Tables 3.3 and 3.4 summarize the results of two of these. These results are presented here for illustration only. It is important to remember that breeds change over time as a result of within-breed selection. So, the results of breed comparisons can become outdated fairly quickly, particularly if some breeds are pursuing effective improvement programmes and others are not, or if breeds are selecting for different characteristics.

There are many recent examples of selection between breeds in the livestock industries. In Britain, many dairy herds of Shorthorn, Ayrshire and Channel Island breeds (Jersey and Guernsey) changed to Friesians in the 1950s and 1960s. In the 1980s and 1990s, the local strains of black-and-white dairy cattle in many European and other countries were partly or wholly substituted by North American Holstein strains (see Fig. 3.5). (Most of these different black-and-white strains originate from importations of Dutch Black Pied animals at some time in the last few hundred years, but they have been selected for different objectives, with different rates of progress in different countries, since then.) Typically, the breed substitutions have been prompted by the higher total milk yield of the incoming breed or strain. In many temperate beef-producing countries the traditional British beef breeds have been replaced, at least as terminal sires, by the larger, leaner, continental European breeds such as the Charolais, Limousin and Simmental. In the British sheep industry, many rams from the Border Leicester breed have been replaced by rams from the more prolific Bluefaced Leicester breed for mating to draft hill ewes to produce crossbred breeding females for commercial flocks. Over the last few decades, some commercial dam lines of pigs have been influenced by importations of the prolific Meishan breed from China, while many sire lines have included importations from the Duroc and Pietrain breeds because of their carcass and meat attributes. Table 3.5 shows examples of breed introductions in Africa and South/South-East Asia.

Table 3.3. Results of a comparison between different sire breeds of sheep, mated to Mule (Bluefaced Leicester x Swaledale) ewes. These results are from just a sample of the breeds involved in a large trial run by the Meat and Livestock Commission in Britain. The results here are from 'late flocks' producing lambs off grass, forage crops and roots. The sire breed means are for lambs at the same estimated level of subcutaneous fatness. The superscripts show whether or not sire breeds differ significantly. Within a row, sire breeds with different superscripts differ significantly; those with the same superscript do not differ significantly, e.g. Oxford Down cross lambs were significantly older at slaughter than Southdown crosses, but were not significantly different in age from Suffolk or Texel crosses. (After Kempster et al., 1987.)

Trait	Sire breed mean			
	Oxford Down	Southdown	Suffolk	Texel
Age at slaughter (days)	258 ^c	214 ^{ab}	228 ^{bc}	262 ^c
Carcass weight (kg)	19.7 ^c	17.0 ^a	18.6 ^b	19.2 ^c
Daily gain of carcass lean (g)	46 ^{bc}	50 ^c	49 ^c	46 ^{bc}
Daily gain of carcass fat (g)	23 ^{bc}	25 ^{cd}	23 ^{bc}	20 ^{ab}

Table 3.4. Results of a comparison between different sire breeds of beef cattle, mated to Hereford x Friesian and Blue Grey suckler cows. These results are from just a sample of the breeds involved in a large trial run by the Meat and Livestock Commission in Britain. The results here are from winter fattening systems. The sire breed means are for steers at the same estimated level of subcutaneous fatness. The superscripts show whether or not sire breeds differ significantly. Within a row, sire breeds with different superscripts differ significantly; those with the same superscript do not differ significantly, e.g. Charolais crosses were older at slaughter than Aberdeen Angus or Hereford crosses, but were not significantly older than Limousin or Simmental crosses. (After Kempster et al., 1982; Southgate et al., 1982.)

Trait	Sire breed mean				
	Aberdeen Angus	Charolais	Hereford	Limousin	Simmental
Age at slaughter (days)	477 ^c	520 ^a	492 ^b	517 ^a	517 ^a
Carcass weight (kg)	205 ^e	268 ^a	214 ^{de}	247 ^c	258 ^b
Conformation (15-point scale)	9.9 ^b	11.2 ^a	8.7 ^{cd}	11.0 ^a	9.9 ^b
Saleable meat in carcass (g/kg)	725 ^{bc}	727 ^b	719 ^{cd}	733 ^a	720 ^{cd}
Fat trim in carcass (g/kg)	96 ^{abc}	90 ^c	97 ^{abc}	92 ^{bc}	93 ^{abc}



Fig. 3.5. North American Holstein Friesians have had a dramatic impact on the populations of dairy cattle in most temperate countries over the last few decades. (Available at: <https://pixabay.com/photos/holstein-cattle-cows-heifers-field-2318436/>, accessed 8 December 2019.)

Selection Within Breeds or Breed Types

Selection within breeds or breed types involves comparing animals of the same breed type and mating the preferred animals to produce the next generation. This process is usually repeated each generation, and as long as there is genetic variation in the characters under selection, this produces changes in successive generations (compared to between-breed selection, where the changes occur only once per breed substitution).

In any population of farm (or wild) animals, the size of the population can only be maintained if adults produce enough offspring in their lifetime to replace themselves when they die or are culled. The capacity to produce potential replacements varies between species and between sexes. For instance, pigs, poultry and especially fish, produce far more potential replacements per adult than cattle, sheep and goats, and in all species there are generally far fewer replacement males needed than females, since each male is usually mated to several or many females. It is this production of more replacements than needed to maintain population size which provides the opportunity for selection within breeds.

Whenever a group of animals is reproductively isolated, and some animals are allowed to breed while others are not, then within-breed selection is being practised. When the choice of animals to breed is based on their performance or other characteristics, this is also known as *assortative mating*. In livestock breeding, this is usually *positive assortative mating*, i.e. animals are chosen for breeding based on them having similar characteristics – typically their performance in one or more traits of economic or other importance. The phenomenon of positive assortative mating is often observed in humans and wild animals too. (Negative assortative mating, where parents show dissimilar characteristics, is rarer.)

As discussed in Chapter 1, assortative mating and within-breed selection have occurred for many centuries during and following the creation of distinct breeds. However, the use of within-breed selection as a tool for genetic improvement became widely recognized following its successful application in improving racehorses in the 17th century and farm livestock breeds in the 18th century.

It was common in those times to mate closely related animals in an attempt to ‘concentrate’ desirable characteristics in the offspring. For example, strains of animals would be created from a series of father–daughter or mother–son matings. This practice of mating close relatives, called linebreeding, is much less common in livestock breeding today. Linebreeding is an extreme form of *inbreeding* – the mating of animals which have one or more ancestors in common. Eventually inbreeding occurs whenever selection is practised in a closed population, but steps are usually taken to limit it by deliberately avoiding matings between close relatives. Related animals have more genes in common than unrelated animals, and the closer the relationship, the more genes they have in common. This means that animals of outstanding genetic merit, for any characteristic, are likely to have relatives which have higher merit than average. For these reasons, it appears entirely logical to breed from closely related animals, but there is a ‘downside’. As well as having more favourable genes in common, related animals also have more unfavourable genes in common, on average. For instance, there are many recessive genes which cause genetic diseases, or adversely affect reproduction, survival or the overall ‘fitness’ of animals. Because these genes generally act recessively, they only cause problems in animals that carry two copies of the gene (homozygous recessive animals). Across a breed as a whole, there may be very few animals which are homozygous for these particular genes, but the genes are still there, undetected, in heterozygous or

carrier animals. Matings between related animals are more likely to produce offspring which are homozygous for some of these genes, than matings between unrelated animals. So, in some cases breeders may be lucky, and they may only see the benefits of mating related animals of high genetic merit. However, in other cases there may be very unfavourable consequences because of infertility or genetic disease in the offspring.

Today we tend to hear only about the past success stories which resulted from linebreeding or close inbreeding, but for each of these there were probably several failures. In plant breeding, and to a lesser extent in poultry breeding, breeders can afford to try linebreeding or close inbreeding in many different lines, and only continue to breed from those which show favourable responses. However, in other livestock species, because of the longer *generation intervals*, the lower reproductive rate, and the higher value of the individual breeding animals, it is not practical to keep many different lines, and the risks from linebreeding usually outweigh the potential benefits. A more gradual, but inevitable and cumulative increase in inbreeding occurs whenever selection is practised in a closed population (e.g. a breed or a closed herd or flock) over a long period of time. This, too, can result in the same unfavourable effects as linebreeding. The effects of inbreeding, and methods of controlling inbreeding in selection programmes are discussed further in Chapter 4.

Requirements for within-breed selection

The choice of animals to breed from may be based solely on their appearance, or on a subjective assessment of their own, or their relatives' performance. For example, selection may be based on visual appraisal of the breed characteristics of an animal – its shape, colour, etc. – or on subjective estimates of the amount of wool or milk produced, or the size of animals. Alternatively, objective records of the performance of the animal itself, or its relatives, may be used. That is, actually weighing the fleeces or the animals, or measuring the volume of milk produced rather than guessing the production. Often a combination of objective and subjective information is used. Objective methods of within-breed selection (i.e. those that rely on recording performance, rather than making subjective judgements about the merit of animals) have become widely used in pig, poultry and dairy cattle breeding since the 1960s. They have been used to a lesser extent in fish, sheep, goat and beef cattle breeding until recently, but this situation is changing in many countries.

Objective selection within breeds or strains is intended to increase the average level of additive genetic merit (termed breeding value) of the population. Ideally, the steps involved are:

1. Deciding what to improve – termed the breeding goal or *selection objective*. For example, a simple breeding goal could be to increase daily weight gain, while a more complicated breeding goal could be to increase lean content of a carcass. As mentioned before, it is sensible to identify the breed or cross which has the highest merit in this trait or combination of traits, and to decide on the most appropriate breeding strategy (e.g. pure or crossbreeding) before embarking on within-breed selection in the best breed, or in component breeds of the best cross. Also, there must be genetic variation in this trait or combination of traits within the breed if any improvement is to be made – in

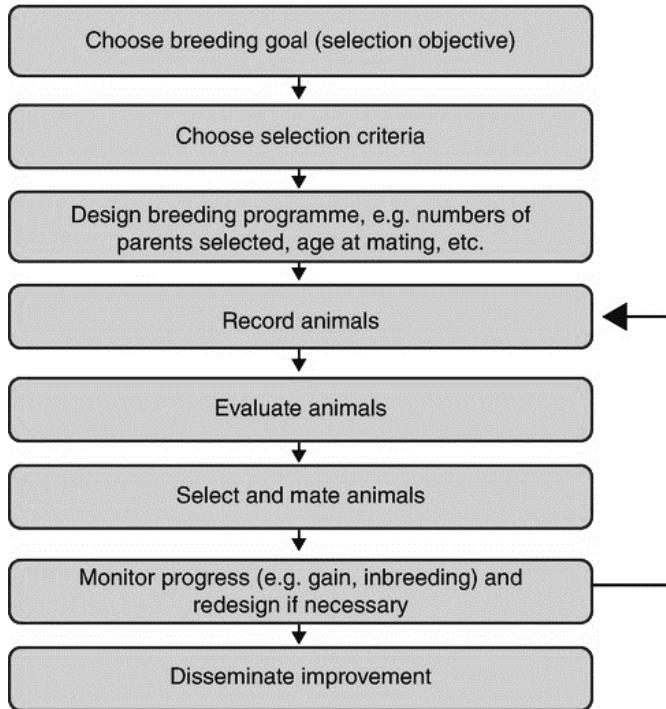


Fig. 3.6. A diagram showing the steps ideally involved in a within-breed improvement programme based on objective measurement of performance. Once the breeding goal is established, different breeds and crosses should be compared in this character, and then within-breed improvement should be used in the best breed, or in component breeds of the best cross. (After Harris et al., 1984.)

other words, there need to be differences between animals in each of these traits, and at least part of these differences need to be inherited.

2. Deciding what to measure and select on within the breed (the *selection criterion*) in order to make improvements in the breeding goal. In some cases, the selection criterion or criteria may be the same as the goal; in others some indirect measurement is needed, for example when the goal trait can only be measured in one sex or after slaughter.
3. Designing the breeding programme (e.g. numbers of males and females to be selected annually; ages at mating).
4. Implementing the programme, i.e. doing the routine recording, evaluation and mating of animals.
5. Monitoring progress and redesigning the programme if necessary.

In most circumstances, unique and permanent identification of individual animals is a requirement for successful within-breed selection. This may sound obvious, but it is harder to achieve than it first seems – particularly when the identities of animals have to be unique across herds, flocks, years or countries. The small size of juvenile fish make this particularly challenging. There are several reasons why unique identification is important. Firstly, it is important to be able to assign records of performance to the right animal, each time its performance is recorded. Secondly, modern methods of *genetic evaluation* make full use of information from relatives, so it is important to be able to establish how recorded animals are related to each other. Thirdly, having recorded and evaluated animals and decided which ones to use for breeding, it is important to be able to find them again! In some special cases, there is justification for simplified recording of identification. For example, in some extensive production systems, animals may be identified only as members of a particular sire family, rather than individually identified. Molecular genetic techniques now allow retrospective identification of parents where identification of young animals is impractical – such as in some

extensive sheep systems or in fish farming. This is achieved using DNA samples from the animals of interest and all of the possible parents to identify the most likely parents.

Objective selection depends on having records of performance on the candidates for selection, or their relatives, or both. In principle, there is no reason why each breeder should not devise their own method for recording performance in traits relevant to their own breeding goal. Most major pig, poultry and fish breeding companies do this. However, in practice cattle, sheep and goat breeders often use recording schemes which are operated by regional or national agencies that specialize in recording and evaluation. (The difference between the species has probably arisen for a number of reasons. These include the relatively small number of pig, poultry and fish breeders in the elite tier compared to the situation in cattle, sheep and goat breeding, so there are fewer ‘products’ for commercial producers to be aware of. Also, many pig and poultry companies aim to fulfil both elite and multiplier rôles, so the breeding companies can communicate directly with their final customers and there is less need for a nationally recognized recording system.) For example, in many countries there are milk recording agencies which visit dairy farms at regular intervals to record the yield and other traits of individual cows, and to sample the milk for analysis of fat and protein content. This information is used to assist management decisions and, when accumulated and analysed, it is used to provide estimates of the genetic merit of cows in the herd and their sires. In the beef and sheep sectors, this performance-recording rôle may be fulfilled by government or industry-funded agencies, breed societies or by breeders themselves. In many countries, the recording agencies adopt the recognized international standards of the *International Committee on Animal Recording* (ICAR) for recording and evaluation methods.

There are several advantages to using regional or national systems in species, like cattle and sheep, with many dispersed breeding flocks or herds:

- they encourage discipline and consistency in which traits are recorded, and how and when they are recorded;
- they often provide some type of authentication of the records of performance – either by supervising some of the recordings or, more commonly now, simply by checking that performance records fall into expected ranges, and that details from different sources (e.g. breeder and breed society) match;
- they usually allow access to more sophisticated methods of data storage, processing and evaluation than is feasible on most farms;
- depending on the method of evaluation and the population structure, they can allow direct comparison of the estimated genetic merit of animals across many herds or flocks – this increases the pool of animals available for selection by breeders, and also allows customers to identify individual animals, or flocks and herds of high genetic merit; and
- there are often benefits when it comes to selling breeding stock, because large recording schemes can achieve a stronger ‘identity’ and better communication about their purposes and products than individual breeders could achieve.

There are also disadvantages to membership of large regional or national recording schemes. There is usually a greater (apparent) cost than with ‘home-grown’ schemes, and often there is less flexibility to tailor breeding goals to individual breeders’ requirements. However, on balance, the advantages to co-ordinated regional or national schemes usually far outweigh the disadvantages.

Factors affecting rates of improvement

As discussed in Chapter 2, many traits of economic or other importance in farm livestock are under the control of a large number of genes and show continuous variation. The response to selection (R) *per generation* in these polygenic traits depends on the *selection differential* (S) and the heritability of the trait under selection (h^2) (after Lush, 1945):

$$R = S \times h^2$$

This is known as the ‘breeder’s equation’, attributable to Jay Lush in the 1930s, one of the founders of animal breeding theory.

The selection differential is the difference between the mean performance of selected animals (i.e. those identified to become parents of the next generation) and the overall mean of the group of animals from which they were selected. This itself depends on two factors – the amount of phenotypic variation in the traits concerned and the *selection intensity* achieved. This is related to the proportion of animals selected to become parents, based on their own performance or that of their relatives. The lower the proportion of animals selected, the higher the selection intensity and hence the better the selected animals will be, on average.

The heritability can also be defined as the proportion of superiority of parents in a trait (i.e. the proportion of the selection differential) which, on average, is passed on to offspring.

Usually, it is helpful to be able to predict *annual* rates of response to selection, and this is achieved by dividing response per generation by the generation interval, so:

$$R(\text{per annum}) = \frac{S \times h^2}{L}$$

The generation interval is the average age of parents when their offspring are born. Essentially, this regulates the speed with which selected animals contribute their better genes to the flock or herd, via their offspring.

These factors are discussed in more detail in Chapter 4. Generally speaking, the higher the selection differential and heritability, and the lower the generation interval, the higher the annual rate of genetic improvement.

The main opportunities for breeders to accelerate rates of improvement are through choice of the most accurate methods of predicting breeding values – accuracy also depends on the heritability of the trait of interest and the amount of information available on the animal and its relatives – and by maintaining high selection intensities and low generation intervals. However, there are biological limits to the extent to which selection intensity and generation interval can be altered. It is possible to achieve much higher selection intensities in species or breeds with a high reproductive rate. Similarly, shorter generation intervals can be achieved in those species or breeds which reach sexual maturity at a younger age. Largely because of biological advantages in these reproductive characteristics, higher rates of genetic change are possible in fish, pigs and poultry than in ruminants (see Fig. 3.7). However, the reproductive rate of female cattle and sheep can be altered dramatically

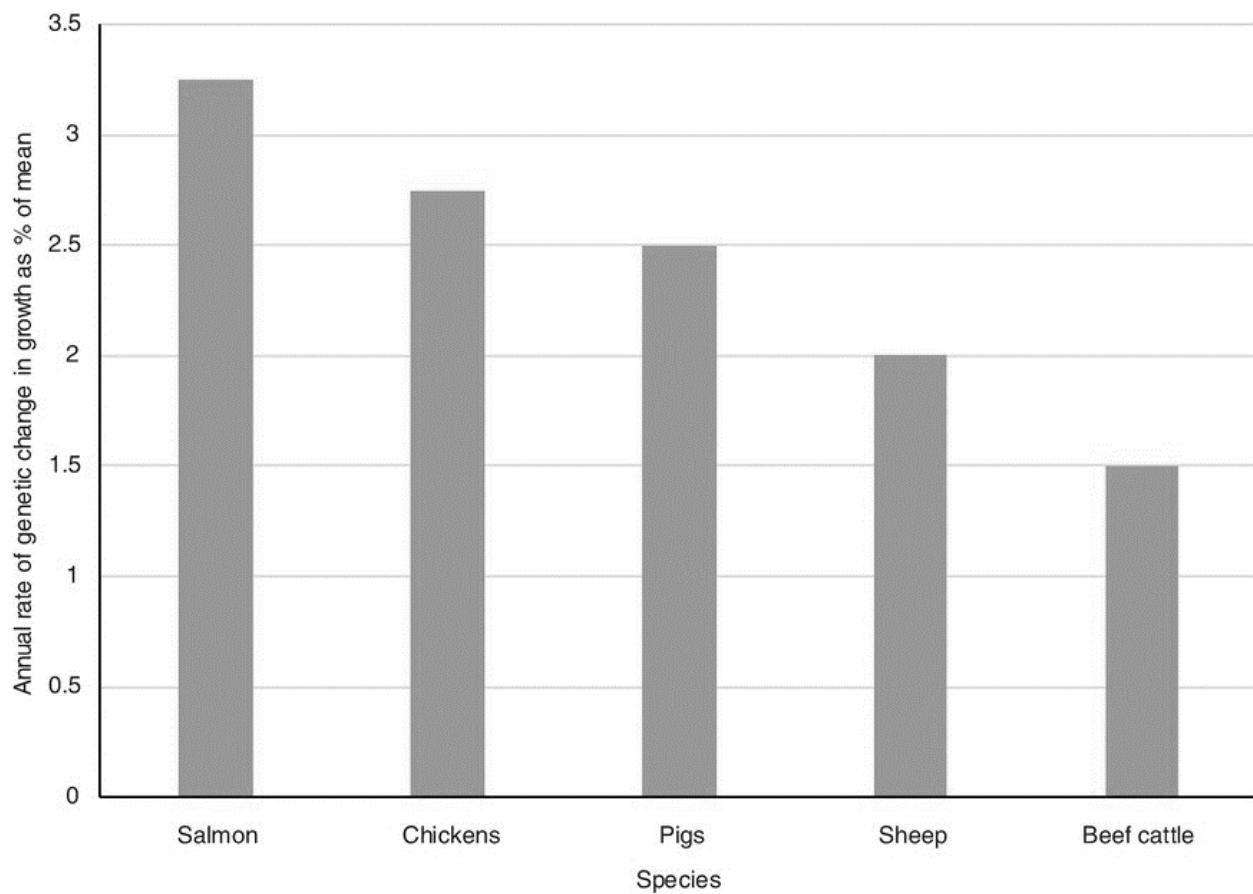


Fig. 3.7. Annual rates of genetic change possible in growth rate in different livestock species. The rates of genetic change are expressed in percentage units of the average growth rate of the species concerned. The differences in rates of change between species are largely due to differences in the natural reproductive characteristics of the species. NB this is for illustration only – few modern breeding programmes select for a single characteristic. (After Smith, 1984 and Simm, 1998; salmon data from Gjedrem and Rye, 2018.)

(at a cost) by the use of reproductive technologies such as *multiple ovulation and embryo transfer*. This effectively makes cattle and sheep ‘reproductively’ more like poultry and pigs and enables higher rates of genetic gain. The use of these technologies is discussed in more detail in Chapter 5, and species-specific examples are given in Chapters 8 to 13.

Selection for more than one trait

In most livestock production systems profitability depends on several different animal characteristics rather than any single trait. For instance, in dairy cattle income often depends not only on yield, but also on fat and protein content of the milk, and often on somatic cell count too. Feed, health and rebreeding costs are important variable costs. In meat animals, returns per animal depend not only on carcass weight, but also on measures of carcass quality such as fatness and conformation, and on feed costs. In genetic improvement programmes it is important to reflect the fact that several traits

influence profit, and so animals are usually selected (whether objectively or subjectively) on a combination of traits. This can be achieved in a number of ways.

One approach, called *tandem selection*, involves selection for one trait for one or more generations, followed by selection for a second trait for one or more generations, possibly followed by selection on more traits, eventually returning to selection on the first, and so on. It can be useful in some circumstances – for instance, if there is an urgent need to improve one trait and other traits need only very minor improvement, it might be appropriate to select only on the first trait until some economically important threshold was reached, and then to turn to the others. However, if the associations between traits of importance are unfavourable, the selection goes round in circles, improving one trait in one generation, and then wholly or partly cancelling that improvement in the next generation when selection is for a second trait.

A more consistent method, called *independent culling levels*, is to set minimum qualifying standards, or thresholds, in several traits of interest. To qualify for selection, animals must then surpass the qualifying standard in each trait. There are methods available to allow the most appropriate qualifying standards, or cut-off points, to be calculated. Figure 3.8(a) illustrates this method of selection. In this example beef bulls are only selected if they achieve a 400-day weight above 560 kg, and a muscling score of over 12 points (on a scale of 1 to 15). An informal version of this approach is used by most breeders when they check candidates for selection for breed type or any ‘functional’ defects (such as overshot or undershot jaws, locomotion faults or small testicles) before breeding from them. This approach is probably a very efficient way of dealing with functional defects which occur irregularly but are suspected of having a genetic component, or traits of minor importance. A version of independent culling levels, or sequential or *two-stage selection*, may also be useful when one trait of interest is very difficult or expensive to measure. For example, it may not be cost effective to record and select on feed intake or feed efficiency on all animals in a selection programme, but it may well be effective to do so, say, on those with weaning weights in the top 25% of those tested.

In theory, the optimal method of selection when there are several traits of economic or other importance is to calculate a selection index. This is a score of overall genetic merit for each of the animals available for selection, based on their own or their relatives’ performance in the traits of interest. In some respects, *index selection* is similar to independent culling levels – animals are selected for more than one trait simultaneously, and altering the emphasis on a trait in a selection index is broadly similar to moving the cut-off point for culling with independent culling levels. However, with index selection an animal can compensate for poor performance in one trait by excelling in another. Although the individual animals that are selected may not have the ideal combination of characters, this is the most efficient way to move a whole population in the desired direction. Index selection is illustrated in Fig. 3.8(b).

Table 3.7 gives a simple example of how an index can help in selection. The table shows milk fat and protein yields for three dairy cows. If selection was based solely on protein yield, then cow A would be selected. If selection was based solely on fat yield, then cow B would be selected. However, if selection was based on fat plus protein yield – a simple index – then cow C would be selected. In this case, the index gives equal emphasis to protein and fat yields, so the two are simply added together. If the desired emphasis on protein:fat was 2:1, then animals could be selected on twice their protein yield, plus their fat yield, and so on. In practice, the traits of interest are often measured

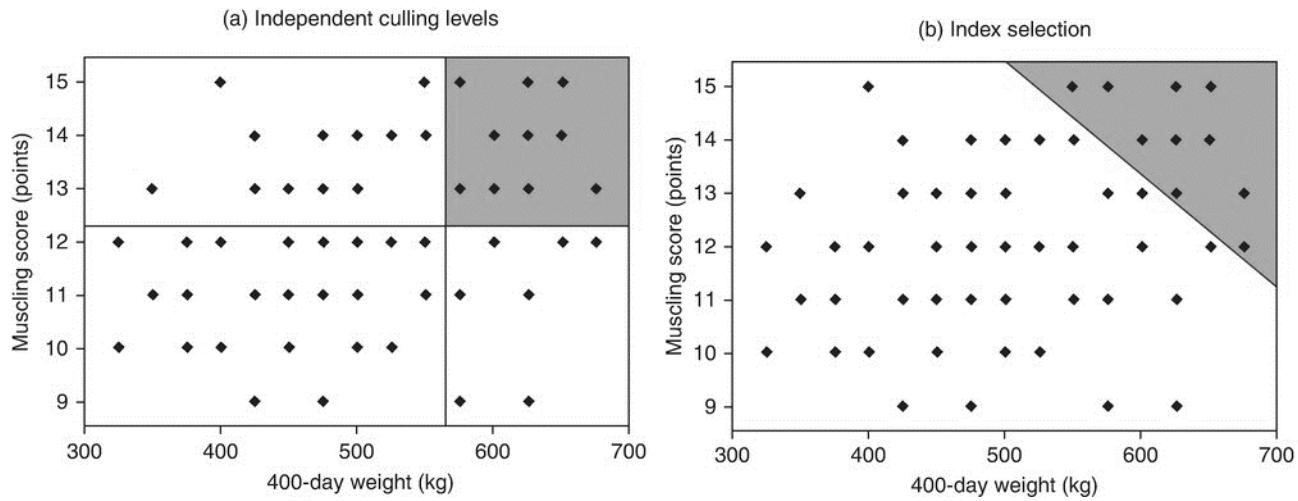


Fig. 3.8. The use of independent culling levels or index selection to select simultaneously for increased 400-day weight and increased muscling score. Each point on the two graphs shows the 400-day weight and muscling score of a single beef bull. (a) Independent culling levels involves setting minimum qualifying standards, or thresholds, in each trait of interest. To qualify for selection, animals must then surpass the qualifying standard in each trait, represented by the vertical and horizontal lines on the graph. In this example, bulls are only selected if they achieve a 400-day weight above 560 kg and a muscling score greater than 12 points (on a scale of 1–15). (b) A selection index scores animals based on all traits of interest, but there are no fixed cut-off points for individual traits – an animal can be selected with lower performance in one trait, providing that it excels in another. Animals above the diagonal line in the graph would be selected. In this example, eight of the ten selected animals would be the same whichever method was used; two of the ten selected animals would differ, depending on the method used.

in different units, and they have different heritabilities, and so arriving at the most appropriate weighting factors is a bit more complicated than it seems at first sight. To derive these weighting factors, it is necessary to know: (i) how much additive genetic variation there is in the traits of interest; (ii) the direction and strength of associations among these traits; and (iii) their relative economic importance. Obtaining reliable estimates of the genetic variation in traits and associations among them requires comprehensive recording of hundreds or thousands of animals. When the traits of interest are already recorded in a large regional or national scheme then there should be plenty of records to obtain these estimates. However, it will be more difficult and costly to obtain them for traits which are not already recorded, which can be a major limitation to the wider use of index selection. Selection indexes are discussed further in Chapter 4 and applications are given in Chapters 8 to 13.

Cow	Protein yield (kg)	Fat yield (kg)	Protein + fat yield (kg)
A	230	280	510
B	200	300	500
C	225	295	520

Table 3.7. An example of the use of a simple index to select cows on protein plus fat yield. The yields in bold type show which cow would be selected if selection was for protein yield only, fat yield only, or protein plus fat yield.

Crossbreeding

Reasons for crossing

Crossbreeding involves mating animals of different lines, breeds, breed types or species. It is used usually for one or more of the following reasons:

- **To improve the overall efficiency of a production system by crossing breeds or breed types which have high genetic merit in different traits (*complementarity*).** The use of specialized sire and dam breeds or lines is a good example of this (Smith, 1964). It is common for commercial herds and flocks in all of the terrestrial meat-producing species to be made up of breeding females from small- or medium-sized breeds or crosses, which have relatively low maintenance costs, and good reproductive and maternal characteristics. These are mated to males from larger terminal sire breeds, with faster growth rate and better carcass characteristics. The use of different breed types with complementary characteristics usually results in production systems with much higher overall efficiency and profitability than those based entirely on small breeds with good reproductive and maternal traits, or those based entirely on large breeds with good growth and carcass traits (Dickerson, 1978).
- **To produce individual animals of intermediate performance between that of two more extreme parent breeds or breed types.** At first sight this is similar to the use of crossbreeding to exploit complementarity. However, the emphasis here is on creating individual animals of intermediate performance, rather than matching different breeds with different rôles in a crossing system. Beef × dairy suckler cows, such as Hereford or Angus × Holstein Friesians, provide a good example of this use of crossing – the crossbred cows have higher beef merit than pure dairy cows, but higher milk yield than pure beef cows. Also, beef × beef crosses, such as the Blue Grey, combine the hardiness of the Galloway with the faster growth and likely higher milk yield of the Shorthorn. In some smallhold livestock production systems in developing countries (e.g. dairy cattle, pigs), it is becoming more common to cross indigenous breeds (that are typically well adapted to the local environment conditions but have low productivity) to improved exotic breeds (that are typically poorly adapted to the local environmental conditions but highly productive), to create animals that are intermediate in terms of both adaptation and productivity (Marshall *et al.*, 2016).
- **For grading up to a new breed or breed type.** As explained earlier, this involves mating animals (usually females) of an existing breed type to those (usually males) from a new breed type (see Table 3.2). The progeny from these matings are themselves mated to the new breed or strain, and this process continues for several generations. For example, the spread of North American Holsteins to many temperate and some tropical countries has been brought about largely by grading up indigenous populations of black-and-white (or other) dairy or dual-purpose cattle.

- **As an intermediate step in the creation of a new synthetic or composite breed type.** Creating a synthetic breed or line usually involves crossing two or more different breeds or breed types. If more than two breeds are involved, then this step can take several generations. For instance, if four breeds are involved (A,B,C and D), two different crossbred types could be produced in the first generation (AB and CD). In the second generation, these two crossbred types could be mated to each other, to produce offspring which have a quarter of their genes from each of the four original breeds. Males and females of these four-way crosses could then be mated to each other in subsequent generations.

When two breeds are crossed, the offspring (called the F₁ generation) are relatively uniform, since exactly half of each animal's genes come from each parent breed. However, if F₁ males are mated to F₁ females (i.e. if the F₁ generation is *interbred*), the resulting offspring, called the F₂ generation, show huge variation in appearance and performance. This is because segregation and recombination have now occurred, leading to a wide variation in the proportion of genes which individual animals inherit from the original breeds (as explained in Chapter 2). For any particular characteristic, some offspring will resemble one of the original parent breeds, while others will resemble the other parent breed, and most will be somewhere in the middle. Following this explosion of variation in the F₂ generation, it takes several generations of selection for the desired characteristics before the variation in performance and appearance is reduced, and a recognizable breed type emerges.

Most synthetic lines of pigs and poultry have been created in this way, although they are not regarded as new breeds since, in most cases, these lines are still open to the introduction of yet more breeds or other synthetic lines. However, there are several examples of synthetic breeds in cattle and sheep. The Luing cattle breed was created from crosses between Shorthorn and Highland cattle, subsequently interbred. There are several synthetic dairy cattle breeds in the tropics which have been created by crossing more productive *Bos taurus* dairy breeds with indigenous heat- and disease-tolerant *Bos indicus* breeds (e.g. the Australian milking Zebu, formed from the Sahiwal, Red Sindi and Jersey breeds and the Jamaica Hope formed from the Jersey, Friesian and Sahiwal breeds (Wiener, 1994)). The Coopworth dual-purpose sheep breed from New Zealand was created by crossing Border Leicester and Romney sheep, followed by interbreeding the crosses with intense selection for numbers of lambs born or weaned, weaning weight, fleece weight and quality and 'easy care' characters, such as lambing ease (Beatson, 1993; see Fig. 3.9). The Meatlinc is a synthetic terminal sire sheep breed, created in Britain by crossing animals from several breeds, including the Suffolk, Dorset Down, Ile de France, Berrichon du Cher and Charollais, and selecting for growth and carcass traits (Fell, 1979). As mentioned earlier, there has been a recent growth in interest in the use of composite beef and sheep breeds, to better meet changing market requirements (see, for example, Innovis (2019); Kelso (2019); Leachman (2019)). There are also examples of 'open composites' with ongoing introduction of new breeds, together with *ad hoc* crossing systems, that share some of the benefits of creating more formal composites, in both higher and lower-income countries (e.g. open beef composites in N America; ongoing *ad hoc* crossing among local Zebu × imported *Bos taurus* crosses in some African smallholder dairy systems; crossing of composite Red Dorper sheep (derived from the Dorset Horn and the Blackhead Persian) with local breeds in E Africa).

Fig. 3.9. Coopworth ewes near Wanganui, New Zealand. The Coopworth is a synthetic breed, the development of which has been based on objective performance recording.



- **To introduce new variation to numerically small breeds or breed types.** In many numerically- small breeds it is difficult for breeders to find enough unrelated animals of sufficiently high genetic merit to sustain a genetic improvement programme. Often, these problems are exacerbated because the population is inbred and genetic defects have emerged which need to be controlled. In these cases, it is quite common for animals from other breeds to be introduced, either officially or unofficially! This is similar to the creation of a synthetic breed, except that the long term aim is to have a much lower proportion of genes from the new breed.
- **To introduce a single gene for a favourable characteristic to an existing breed or breed type (*introgression*).** Crossing is sometimes used to introduce a single gene for a favourable characteristic to an existing breed. Good examples include the introgression of the gene controlling polledness into naturally horned cattle breeds, and the introgression of the Booroola gene affecting fecundity, originally found in Merino sheep, into other sheep breeds. Unlike the creation of a synthetic breed, the aim of introgression is usually to introduce and retain only the desired gene from the new breed, and to remove the other genes contributed by this breed by successive generations of backcrossing to the original breed. It is important in this case to ensure that as many as possible of the females retained for backcrossing carry the desirable gene. After several generations of backcrossing, carrier animals can be mated to each other, to create offspring which are homozygous for the newly introduced gene, but in other respects are similar to the original breed. (See Chapter 5 for more details on how molecular genetic techniques can assist in this process.)
- **To exploit heterosis or hybrid vigour.** When two breeds are crossed, intuitively we expect the performance of the crossbred offspring to fall midway between that of the parent breeds (i.e. at the mid-parent mean). However, in practice, the performance of crossbreds is often better than we expect. This advantage in performance above the mid-parent mean is called heterosis or hybrid vigour (see Fig. 3.10). It is measured either in the units in which the trait was originally measured, or as a percentage increase over the mid-parent mean. For example, if two sheep breeds have average litter sizes of 1.0 and 2.0 lambs, respectively, we expect crossbred ewes to have an average litter size of 1.5 (the mid-parent mean). If, in fact, the crosses have an average litter size of 1.6 lambs, this indicates

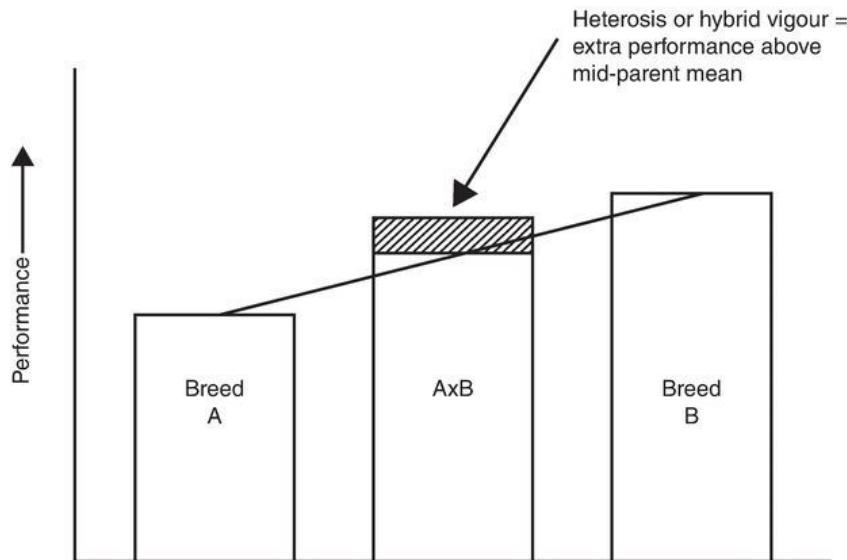


Fig. 3.10. A graph illustrating the phenomenon of heterosis or hybrid vigour. Heterosis is defined as the advantage in performance of crossbred animals above the mid-parent mean of the two parent breeds. In this case the amount of heterosis is the shaded part of the middle bar in the graph. Heterosis is most useful when it leads to the average performance of the crossbred animals exceeding that of the best parent breed in a single important trait or a combination of important traits.

that there is heterosis in litter size of 0.1 lamb ($1.6 - 1.5$) or 6.7% ($0.1/1.5 \times 100$). Heterosis is usually greatest in traits associated with reproduction, survival and overall fitness. (Note that heterosis can sometimes be negative/detrimental as well as positive/beneficial; see, for example, Bunning *et al.* (2018).) There are many important examples of heterosis in animal breeding. The beneficial effects of heterosis occur for exactly the opposite reason that the detrimental effects of inbreeding occur. In other words, when two breeds or lines are crossed there is a much lower proportion of offspring which are homozygous for recessive genes affecting reproduction, survival and overall fitness, or causing genetic disease, than if animals of the same breed are mated. Crossbreeding creates animals which are heterozygous at more loci, whereas *purebreeding* creates animals which are homozygous at more loci.

Selection between breeds, selection within breeds, and the first four applications of crossbreeding listed above, all exploit differences in additive genetic merit between populations or individual animals. However, heterosis is the result of non-additive gene action. That is, it is a result of dominance at individual loci, or epistasis between loci, or both (see Chapter 2 for details). The fact that heterosis is a result of non-additive gene action means that it is difficult to predict the amount of heterosis to expect when particular breeds are crossed. Some crosses result in substantial heterosis and others do not. When a particular cross produces a large amount of heterosis, the parent breeds are sometimes said to *nick well*, or show good *combining ability*. Although it is difficult to predict the level of heterosis that will arise from crossing any two breeds, it is usually greater for crosses between genetically diverse breeds. For example, heterosis is usually greater in beef x dairy cattle crosses than in crosses between two dairy breeds, and it is usually greater in crosses between *Bos taurus* and *Bos indicus* cattle breeds, than in crosses between two European *Bos taurus* or two *Bos indicus* breeds; it appears to be greater still in crosses between tropical and European *Bos taurus* breeds (Bunning *et al.*, 2018). This is probably because the more distantly related the two breeds, the greater the proportion of loci at which different alleles are fixed in the two breeds, and hence the greater the number of loci at which the crossbred offspring are heterozygous. As mentioned above, heterosis is usually greatest for traits affecting reproduction, survival and overall

fitness, and it is usually least for ‘production’ traits like growth and milk yield. Figure 3.11 shows the relative milk yield of indigenous *Bos indicus* cattle breeds, *Bos taurus* dairy breeds, F₁ crosses between indigenous and exotic dairy breeds, and subsequent generations of interbred animals from a review of many dairy cattle crossbreeding experiments in the tropics.

Heterosis is a useful bonus if there are other primary reasons for crossbreeding, for example, to exploit complementarity of breeds. However, crossbreeding solely to exploit heterosis is only really justified if the heterosis is sufficient to make the crossbred animals better, on average, than the best parent breed. In other words, having substantial heterosis is of no net benefit if the crossbred animals are still inferior to purebred animals of one of the parent breeds. So, any evaluation of crossbreeding as a strategy for genetic improvement needs to take into account the additive genetic merit of the pure breeds, as well as the non-additive ‘bonus’ that occurs when they are crossed.

There are numerous examples of experiments to compare both pure breeds and crosses of cattle and sheep. The results from one of these, an experiment to compare the reproductive and maternal performance of three different types of crossbred sheep in Britain, are shown in Table 3.8. Others include a large ‘multi-breed’ experiment, established at the Animal Breeding Research Organisation (ABRO) in Edinburgh in 1970, to estimate the extent of between-breed differences in traits affecting overall efficiency. The experiment involved 25 cattle breeds of different mature size and different levels of milk production. The results showed that at about 1 year of age, between-breed variation accounted for about 70% of the total variation in body weight and 60% of the total variation in cumulated food intake, which highlights the importance of selecting the most appropriate breed (Thiessen *et al.*, 1984). In the 1970s and 1980s, a series of large-scale experiments was done at the US Meat Animal Research Center in Nebraska, to compare the growth, carcass characteristics and reproductive performance of many beef breeds and crosses, to compare crossbreeding systems and to compare different synthetic strains (Cundiff *et al.*, 1982, 1986; Gregory *et al.*, 1982; see Leachman (2019) for industry application of these results). A similar experiment was established in the early 1990s at the Australian Beef Co-operative Research Centre in Armidale, with the emphasis on investigating differences in meat quality between breeds and crosses.

Fig. 3.11. The relative milk yield of indigenous *Bos indicus* cattle breeds, F₁ crosses between *Bos indicus* and *Bos taurus* dairy breeds, *Bos taurus* dairy breeds, and subsequent generations of interbred animals from a review of many dairy cattle crossbreeding experiments in the tropics. Yields are expressed in percentage units relative to that of the *Bos indicus* breeds.
(After Cunningham and Syrstad, 1987; Davis and Arthur, 1994.)

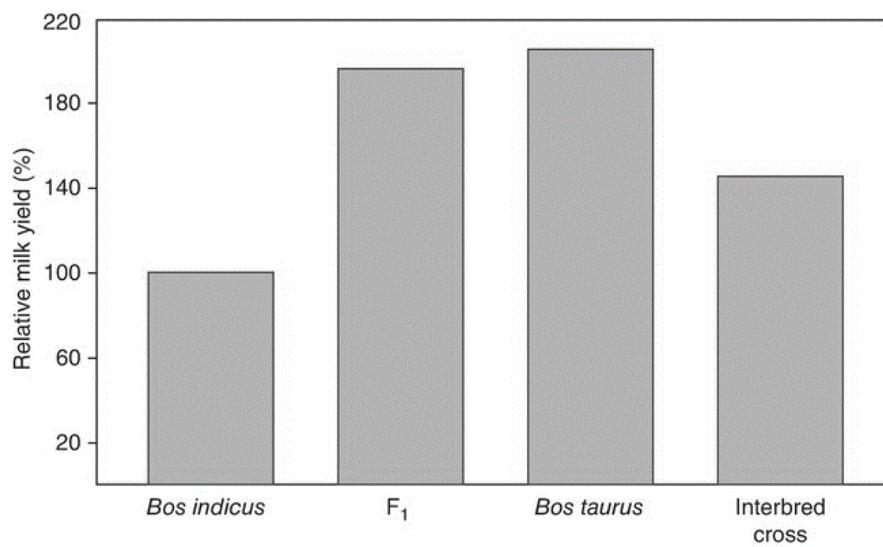


Table 3.8. Results of a comparison between three different crossing breeds of sheep: the Border Leicester, Bluefaced Leicester and ABRO Damline. The table shows the average performance of crossbred ewes produced by mating rams from these three breeds to ewes of several hill breeds. The crossbred ewes were mated in their first year, which partly explains the relatively low levels of performance. Matings were to a terminal sire breed. The superscripts show whether or not the crossbred ewe types differ significantly. Within a row, crossbred means with different superscripts differ significantly; those with the same superscript do not differ significantly, e.g. significantly fewer Border Leicester cross ewes lambed in their first year, compared to the other two crosses, but there was no significant difference in the proportion of Bluefaced Leicester cross or Damline cross ewes lambing in their first year. (After Cameron et al., 1983.)

Trait	Crossbred ewe type		
	Border Leicester ×	Bluefaced Leicester ×	ABRO Damline ×
Ewes lambing per ewe mated in first year of life	0.62 ^a	0.73 ^b	0.78 ^b
Total number of lambs born per ewe mated (avg. of first three matings)	1.29 ^a	1.40 ^b	1.54 ^c
Number of lambs born alive per ewe mated (avg. of first three matings)	1.16 ^a	1.28 ^b	1.37 ^c
Estimated litter weight at 10 weeks, per ewe mated (kg)	25.94	31.10	28.68
Productivity relative to Border Leicester cross (=100) taking into account estimated ewe feed requirements	100	114	117

In most temperate dairy industries, there is very little systematic use of crossbreeding. A major reason for this is that the milk production of crossbreds rarely exceeds that of the pure Holstein Friesian, although there is heterosis for several important traits. In contrast, there is widespread use of systematic crossing, especially between Jerseys and Holstein Friesians, in the extensive pasture-based dairy industries of New Zealand and to a lesser extent Australia. (In 2017/18, Holstein Friesian-Jersey crosses, Holstein Friesians and Jerseys accounted for around 48%, 33% and 9%, respectively of dairy cows in New Zealand (Livestock Improvement Corporation Ltd and DairyNZ Ltd, 2018).) Both of these breeds have a long history of successful within-breed selection in New Zealand, and so have high additive genetic merit for production from grass. Additionally, crossbred animals appear to be particularly favoured when comparisons are made on the basis of milk production per hectare. Since heterosis for production is higher than that for liveweight, crossbred animals achieve the extra production above the parental mean without incurring the ‘penalty’ of higher live weight, and hence higher maintenance costs (Harris et al., 1996).

Types of heterosis

When parents of two different breeds are mated, the crossbred progeny may show heterosis in a range of characteristics. These might include the time taken to get up and suck after birth, neonatal survival and early growth. Once the crossbred animals themselves mature and reproduce, heterosis may be evident in another set of traits associated with fertility and maternal ability. At this stage, some of the benefits of heterosis accrue to the offspring of the crossbred female, rather than to the female herself. So, it is useful to distinguish between:

- *Individual heterosis* – influencing performance as a result of animals themselves being crossbred.
- *Maternal heterosis* – influencing the reproductive and other maternal performance characteristics of crossbred females. The benefits in this case are often measured in the offspring (e.g. extra weight of offspring weaned by crossbred cows, compared to purebreds). Maternal heterosis occurs as a result of dams being crossbred.
- *Paternal heterosis* – influencing the reproductive performance of crossbred males. Paternal heterosis occurs as a result of sires being crossbred.

Although there may be paternal heterosis for traits such as libido and male fertility, individual and maternal heterosis are of most practical value. Table 3.9 gives some examples of individual and maternal heterosis in economically important traits in crosses between breeds of sheep, beef and dairy cattle, pigs and poultry. There are several good examples of industry structures, or breeding schemes, which make good use of both individual and maternal heterosis, as well as complementarity. There is widespread, organized use of crossing in the pig and poultry industries. The aim is to produce breeding females for commercial herds and flocks which, as well as having high additive genetic merit for reproduction and associated characteristics, show maternal heterosis. Mating these females to males of a different breed or strain/line then maximizes individual heterosis in the offspring. For example, in the pig industry in the UK and elsewhere, commercial F₁ females are produced typically by crossing strains based on the Large White and Landrace breeds, which have been selected for high maternal performance. In commercial herds these sows, in turn, are mated to boars from breeds, or synthetic lines based on several breeds, including the Hampshire, Duroc and Pietrain, which have been selected for growth, carcass traits and feed conversion efficiency.

Crossbreeding schemes in sheep and beef cattle have often evolved over time in a less planned way than in the pig and poultry industries. For instance, some production systems based on crossbred animals have probably been stimulated by the availability of relatively cheap F₁ females as a by-product of other enterprises, for example the use of beef × dairy suckler cows in Britain and Ireland (see Fig. 3.12) and the use of longwool × Merino ewes or longwool × hill ewes for lamb production in Australia and the UK, respectively. Nonetheless, the benefits are still obtained.

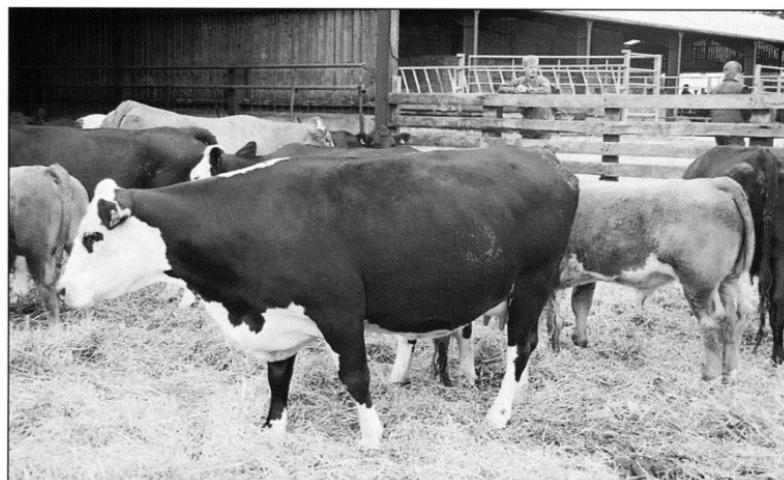


Fig. 3.12. Beef × dairy suckler cows have been very important historically in Britain and Ireland. However increasing specialization in the dairy herd, and reduced dairy cow numbers, means that more beef producers in these countries are breeding their own beef × beef replacements.

Table 3.9. Examples of heterosis in traits of economic importance in livestock. Values shown are specific for the combination of breeds concerned. (After Simm *et al.*, 1994.)

Species/type of animal	Breeds crossed	Trait	Amount of heterosis in units of measurement		Source
			Type of heterosis	or as a % of the mid-parent mean	
Sheep	Galway, Border Leicester, Cheviot and Blackface	Fertility	Maternal	7.3%	Timon (1974)
		Litter size	Maternal	2.3%	
		Lambs/ewe mated	Maternal	10.0%	
		Lamb wt/ewe mated	Maternal	11.8%	
		Lamb mortality	Maternal	2.0%	
	Several breeds	Lamb daily gain	Individual	2–4%	Mayala (1974)
Beef cattle	Hereford, Angus, Shorthorn	Weight calf weaned per cow exposed	Maternal	14.8%	Cundiff <i>et al.</i> (1982)
		Age at puberty	Individual	8.5%	
		Post-weaning gain	Individual	-9.4%	
	Several breeds	Carcass weight	Individual	11.0%	Cundiff <i>et al.</i> (1986)
		Fat thickness	Individual	15.0%	
			Individual	0.1%	
Dairy cattle	Holstein Friesian, Jersey	Milk	Individual	129 litres (3.9%) ^a	Harris <i>et al.</i> (1996)
		Fat	Individual	6.8 kg (4.1%) ^a	
		Protein	Individual	5.0 kg (4.1%) ^a	
		Survival from 1st to 2nd lactation	Individual	4.7%	
		Liveweight	Individual	7.2 kg (1.9%) ^b	
	Pigs	Average daily gain 70–154 days	Individual	7.3%	Cassady <i>et al.</i> (2002)
	Yorkshire, Landrace, Large White, and Chester White	Average daily feed intake	Maternal	1.1%	
		Carcass backfat	Individual	10.8%	
			Maternal	0%	
			Individual	-0.2%	
			Maternal	2.2%	
Chickens	High and low body weight selected lines of White Plymouth Rock	Body weight 266 days	Individual	13%	Williams <i>et al.</i> , 2002
		Age at first egg	Individual	-23%	
		Hen-day ovulations	Individual	34%	

^a% heterosis derived from mean performance of the two pure breeds in 1995/96 (Livestock Improvement, 1996).

^b% heterosis derived from assumed mid-parent mean liveweight of 375 kg.

The *stratified* system of sheep breeding in the UK provides a good example of the use of complementarity and both maternal and individual heterosis (see Fig. 3.13, Chapter 10 and Meat and Livestock Commission, 1988). Most hill sheep flocks in the UK are made up of purebred ewes of traditional hill breeds such as Scottish Blackface, Swaledale and Welsh Mountain. Typically, these ewes are bred pure in the hills for about four lamb crops. Then they are ‘drafted’ to better land on the same farm, or sold to other farms, for crossing – usually with a ram from a longwool breed, such as the Bluefaced Leicester or Border Leicester. The F₁ females resulting from these matings are widely used in commercial flocks in the uplands and lowlands (see Fig. 3.14). They, in turn, are mated to rams from the terminal sire breeds, such as the Suffolk and Texel, to produce lambs for slaughter. The use of F₁ longwool × hill females in commercial flocks makes full use of maternal heterosis. Crossing these to a third breed maximizes the benefits of individual heterosis in the slaughter generation. Additionally, there are benefits from complementarity, because of the use of medium-sized F₁ breeding females with good reproductive and maternal characteristics, but much larger terminal sires with good growth and carcass characteristics.

Similarly, the use of large terminal sire beef breeds in herds of cows made up of crosses between traditional British beef breeds, or crosses between beef and dairy breeds, makes full use of complementarity and both maternal and individual heterosis. Figure 3.15 illustrates the cumulative benefits of individual and maternal heterosis in beef cattle.

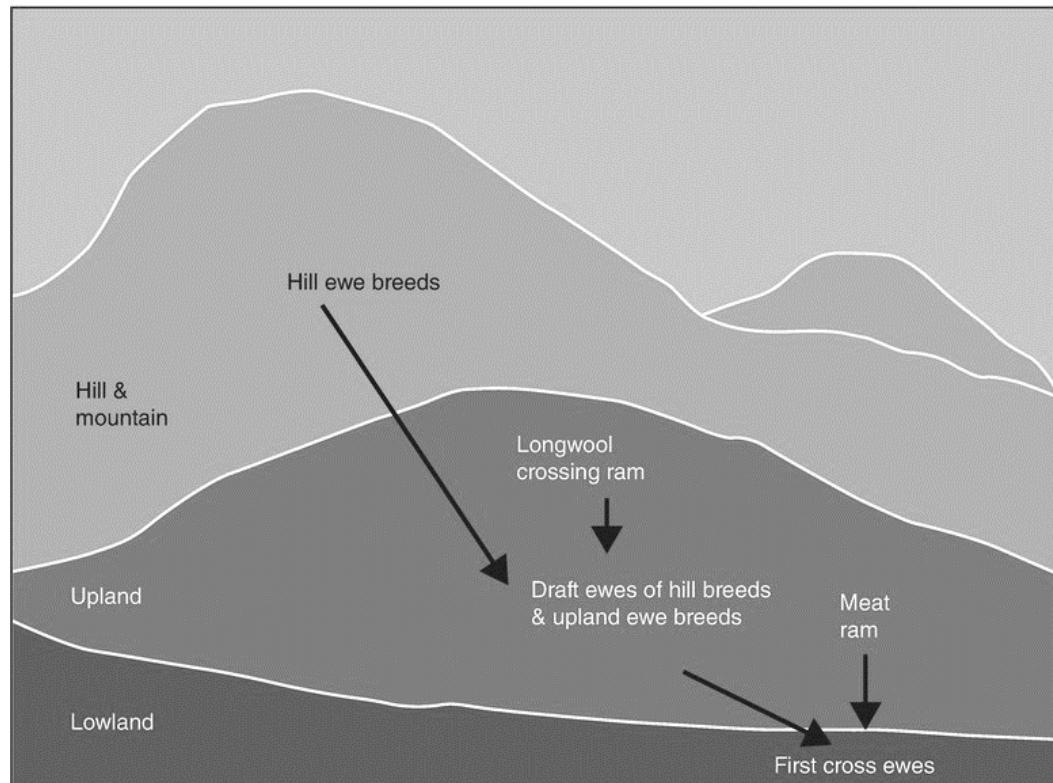


Fig. 3.13. A diagram illustrating the stratified system of sheep breeding in the UK. This system makes use of complementarity of different breeds and of both maternal and individual heterosis. (After Meat and Livestock Commission, 1988.)



Fig. 3.14. Crossbred ewes such as these Scottish Mules (Bluefaced Leicester \times Scottish Blackface) form the backbone of the commercial sheep industry in Britain.

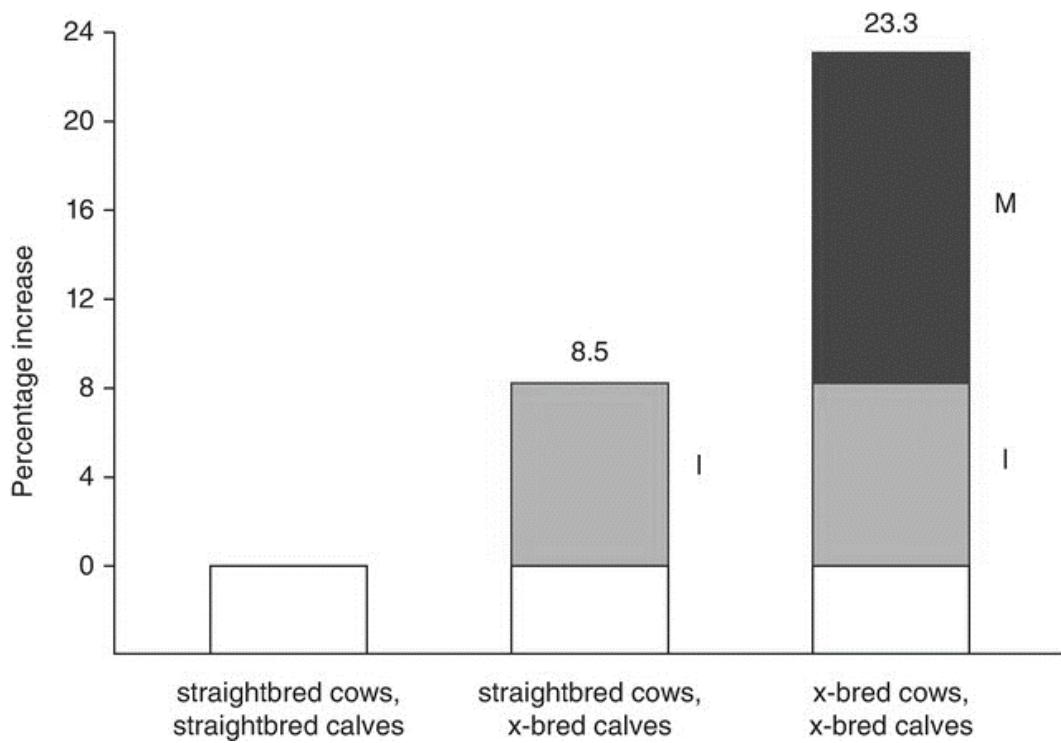


Fig. 3.15. A diagram illustrating the cumulative benefits of individual (I) and maternal (M) heterosis on weight of calf weaned per beef cow exposed to breeding. (After Cundiff and Gregory, 1977; Nicholas, 1987.)

Systems of crossing

The simplest type of cross is that between two breeds. For obvious reasons this is called a *two-way cross*. The progeny resulting from a two-way cross are called F_1 or first cross animals. In the previous section there were several examples of this type of cross, including longwool \times hill breed commercial ewes, and beef \times dairy suckler cows.

If animals from this two-way cross are then mated back to one of the parent breeds, this is termed a *backcross*, as mentioned earlier. When F_1 crossbred animals are mated back to one of the parent breeds, heterozygosity, and so individual heterosis of the resulting offspring, is halved on average, compared to that in the F_1 generation. This is also true in all subsequent generations of backcrossing. Individual heterosis is at its maximum in the F_1 generation and is halved in every subsequent generation of backcrossing to the same parent breed. (This happens because each generation of backcrossing halves the number of loci which are heterozygous, on average). Similarly, when F_1 animals are interbred to produce an F_2 generation, heterosis is expected to halve compared to that in the F_1 generation (though experimental results sometimes depart from this expectation – possibly because of differences in the relative contribution of epistatic compared to dominance effects; see Falconer and Mackay, 1996). No further reduction in heterosis is expected in the F_3 and F_4 generations until inbreeding occurs.

The performance of later generations of crossbred animals, after the F_1 , may also be reduced as a result of *recombination loss* (Dickerson, 1973). This is believed to occur as the result of the breakdown of favourable ‘epistatic blocks’ of alleles from parental breeds, where these blocks have been established by many generations of selection.

Table 3.10 shows the fractions of individual, maternal and paternal heterosis which are expected to be maintained in different types of crossbreeding schemes, *relative to that in the F_1* . For example, a value of 1 in the table indicates that a particular type of cross maintains the full amount of heterosis seen in the F_1 ; a value of $\frac{1}{2}$ indicates that heterosis is halved compared to that in the F_1 . As mentioned before, the value of a particular cross depends on its average merit compared to that of the best parent breed, i.e. it is a function of the additive genetic merit of both of the parent breeds, as well as heterosis.

One way of maintaining heterosis after producing a two-way cross is to mate to a third breed, to produce a three-way cross. The progeny of terminal sire sheep or beef breeds out of the type of F_1 ewes and cows described above are three-way crosses. New breeds can be added to the mix, in an attempt to maintain heterosis, but unless additional breeds have high additive genetic merit, the benefits of keeping heterosis high will soon be outweighed by the use of inferior breeds. Also, although three-way crosses maintain the *relative* level of heterosis compared to that in the F_1 generation, the *absolute* amount will depend on the specific breeds concerned – some combinations of breeds will lead to high absolute levels of heterosis, and other combinations will not.

Systems based on specific crosses can be very efficient, especially if specialized sire and dam breeds are used. However, as illustrated above, in cattle and sheep breeding these systems usually depend on a readily available external supply of replacement F_1 females. Otherwise, substantial numbers of at least one of the constituent breeds will need to be kept alongside the crossbred commercial herd or flock, just to breed sufficient replacement females.

Table 3.10. Fractions of heterosis expected to be maintained in different crossing systems. (After Dickerson, 1974; Nicholas, 1987.)

Type of crossing system (breed of male is given first)	Fraction of heterosis relative to that in the F ₁		
	Individual heterosis (e.g. in survival of the crossbred animal itself)	Maternal heterosis (e.g. in fertility, or total weight of calf weaned by cows)	Paternal heterosis (e.g. in male fertility or libido)
Pure breed	0	0	0
Two-breed cross			
A × B	1	0	0
Backcross			
A or B with AB	1/2	1	0
AB with A or B	1/2	0	1
Three-breed cross			
C × AB	1	1	0
AB × C	1	0	1
Four-breed cross			
AB × CD	1	1	1
Rotational cross			
2 breed	2/3	2/3	0
3 breed	6/7	6/7	0
Composite ^a			
2 breeds, equal contribution	1/2	1/2	1/2
4 breeds, equal contribution	3/4	3/4	3/4
8 breeds, equal contribution	7/8	7/8	7/8

^aRetention (compared to F₁) = approx. $1 - \sum_i^n P_i^2$ where: P_i = proportion of each of n breeds used to create composite (Gregory *et al.*, 1993).

An alternative to the use of specific crosses is *rotational crossing*. This involves the use of the same two or three (or more) breeds in rotation. In a two-breed rotational cross, breeds A and B would be mated to produce F₁ offspring with 50% of the genes of each parent breed (abbreviated to AB here). These would be mated to sire breed A, to produce a second generation (abbreviated A(AB)) of offspring with an average of 3/4 A genes and 1/4 B genes. These, in turn, would be bred to sires from breed B, producing offspring with an average of 3/8 A genes and 5/8 B genes. This process continues until the proportions of genes from the two breeds stabilizes at an average of about 1/3A, 2/3B and 2/3A, 1/3B in successive generations (see Table 3.11). In a three-breed rotational cross the proportions of genes from the three breeds stabilizes at an average of about 1/7, 2/7, 4/7, with the highest proportion of genes coming from the sire breed used to produce the most recent generation, and *vice versa*. In practice, most farmed livestock species live for varying lengths of time, and produce offspring at a range of ages, and so generations overlap rather than being discrete. This means that herds or flocks are soon made up of animals with different proportions of genes from the breeds involved, and several sire breeds need to be used each year. This can lead to additional difficulties in recording, arranging mating groups, and in managing and feeding animals of different genetic make-up if they differ markedly in size or productivity. Two- and three-breed rotational crossing maintains about 2/3 and 6/7, respectively, of the level of heterosis seen in the F₁ generation. However, as mentioned already, it is the combined effect of additive genetic merit and heterosis in the particular crosses or crossbreeding systems concerned that needs to be evaluated.

Rotational crossing is used in the pastoral beef industries of several countries – particularly to breed replacement females which are two- or three-way crosses between the traditional British beef breeds (Aberdeen Angus, Hereford, Shorthorn). There is also interest in this type of system in countries traditionally dependent on beef × dairy replacements, as increasingly specialized dairy breeds, such as Holsteins, usually have poorer beef characteristics than the dairy breeds they replaced.

Composite breeds are sometimes preferred as an alternative to rotational crossing, as they can maintain relatively high levels of heterosis (though lower than with rotational crosses among the same component breeds), while avoiding the variability in performance seen in herds or flocks with different generations of rotational crosses, and reducing complexity of management. For example, composites made up of equal proportions of two, four or eight breeds will retain 50%, 75% and 87.5%, respectively, of the heterosis seen in the F₁. These proportions of heterosis retained will decline as a result of inbreeding in the population.

Some of the long-established cattle and sheep composite breeds mentioned earlier are now treated effectively as pure breeds, in that they are closed to the introduction of new breeds. Often in pig and poultry breeding, and with some of the newer beef and sheep composites, there is a much more fluid concept of the composite, with new breeds or crosses being added if they are expected to enhance overall performance.

Conservation of Genetic Resources

Conservation of animal, plant and microbial genetic resources used in agriculture is vital for future food security and resilience of food systems. All of the strategies for genetic improvement discussed in this chapter depend on genetic variation, so it is important for the success of future improvement programmes that genetic variation is used in a sustainable way. It is useful to consider two types of conservation of genetic variation in domestic animals, although they overlap. The first is the conservation of rare breeds or strains of livestock, which are in danger of extinction. The second is conservation of genetic variation within breeds involved in active improvement programmes. The first of these is discussed below. Conservation of variation within breeds is discussed further in Chapter 4, in the context of designing within-breed improvement programmes.

Why conserve rare breeds?

In several industrialized countries, interest in breed conservation was stimulated from the 1960s onwards by farmers and breed enthusiasts who were concerned that the increasing use of a few specialized breeds was leading to a severe decline in the numbers of animals in many traditional breeds. Organizations like the Rare Breeds Survival Trust (RBST) in the UK and the American Minor Breeds Conservancy (now the American Livestock Breeds Conservancy) were set up to promote conservation of those breeds thought to be at risk. In many lower-income countries, interest in conservation was stimulated, at about the same time, by the widespread crossing of indigenous breeds to imported breeds.

There are a number of economic, social and cultural reasons to conserve rare breeds. Farm animal genetic resources (FAnGR) are the foundation for livestock production, which is responsible for around 40% of the value of agricultural output globally (FAO, 2019). So, there are strong economic arguments to protect FAnGR. Livestock have a particularly important rôle in supporting livelihoods of many of the world's poorest people. Breeds and crosses that can prosper in harsh conditions may make a particularly important contribution to feeding those humans most in need. Rare breeds are reservoirs of genetic variation that are being overlooked and may become important in future, especially in response to the challenges of climate change, food insecurity, and so on. In addition to this utilitarian view, FAnGR are part of wider biodiversity, which many argue has intrinsic value. Also, the wide variety of livestock breeds we have today is part of our cultural heritage and deserves protection for this reason too.

Scientific support for breed conservation has a rather chequered history. The subject received relatively little scientific attention until the late 1900s, perhaps because support for conservation originated largely from 'grassroots' enthusiasts, and bypassed most scientists. However, there is also a counter argument to conservation on biological grounds. This is that the markets for animal products do not usually change dramatically, and so there is scope for popular breeds to keep pace with new markets by changing the emphasis in selection, without the need to return to rare breeds for new genetic variation. There is already wide genetic variation in most breeds, which permits short- and medium-term responses to selection. In the longer term, new variation is created by mutation, and so current breeds should be able to adapt to new markets indefinitely, provided that population sizes are large. While this argument probably holds for breeds that are not that much different from each other in the first place, it is probably not true globally. Many of the breeds or strains most at risk have evolved over a long time in very harsh or specialized environments. It would be difficult and time consuming to reinstate in 'improved' breeds some of the traits, like tolerance to disease, heat and nutritional deprivation, which some breeds at risk already possess. Hence, the case for conservation is widely accepted.

Over the last few decades, the Food and Agriculture Organization of the United Nations (FAO) has had a key rôle in informing international policy and co-ordinating national efforts on conservation and sustainable use of genetic resources. The FAO website (FAO, 2019) has links to many key reports and other resources relevant to conservation of animal genetic resources, including periodic reports on the *State of the World's Animal Genetic Resources*, global action plans and the Domestic Animal Diversity Information System (DAD-IS).

Many breeds or lines of economic or other importance in many countries are imported or have been developed with contributions from foreign breeds or lines. As part of the Convention on Biological Diversity (CBD), *The Nagoya Protocol on Access and Benefits Sharing* is an international agreement intended to ensure that the benefits arising from the utilization of genetic resources, as well as traditional knowledge associated with them, are shared in an equitable manner (see CBD, 2019). Over 116 countries are parties to the Nagoya Protocol. These parties are obliged to establish legislative and/or other measures regarding access to genetic resources. Implementation of the Nagoya Protocol is co-ordinated by an international body known as the Access and Benefits Sharing Clearing House (<https://absch.cbd.int/>).

Which breeds should be conserved?

If the case for conserving breeds is accepted, the next decision is which breeds to conserve. There are around 8800 breeds or strains of livestock globally, belonging to 38 species (FAO, 2019). Seventeen percent of these breeds are believed to be extinct or at risk of extinction, and for another 58% the risk status is unknown. With limited resources to devote to conservation, priorities must be set. International co-operation is important if resources are to be used efficiently. For instance, there is no point in two countries devoting their scarce resources to the conservation of the same or closely related breeds, while others get neglected. (Though the vast majority of breeds are reported in only one country.) Hence, collecting information on the status of different breeds has been a priority for organizations involved in conservation. (See Wainwright *et al.* (2019) for one approach to prioritizing breeds for conservation.)

The FAO DAD-IS system mentioned above (FAO, 2019) collates available information on:

- breeds and strains that exist in each member country;
- whether the breeds exists in other countries (transboundary breeds);
- numbers of breeding animals in each population and risk status;
- conservation programmes in operation;
- genetic and production characteristics of each breed;
- typical production systems; and
- cultural importance.

This information helps in the identification of breeds or strains which are declining in numbers rapidly, though large gaps and errors in the data (as reported by countries) remain. Table 3.12 shows the ‘headline’ threshold population sizes used by FAO to assign risk status to livestock breeds. These are further refined, depending on the species reproductive rate, trends in population size, and extent of pure breeding. Other institutions or countries use higher population thresholds to declare populations at risk, recognizing that it may be more cost effective to act sooner where populations are declining. This is especially true in countries with constrained resources or infrastructure. It is also important to note that breeds may be at risk not just because of numerical scarcity, but also because of low genetic variability, because of geographical concentration or because of adaptation to a specific environment. The UK considers native breeds to be at risk when populations of breeding females fall below 5000 (equines), 7500 (cattle), 10,000 (sheep and goats), 15,000 (pigs) or 25,000 (poultry). For more information on this, on UK National Action Plans and other resources, see Defra (2019).

Historically, the degree of similarity between breeds has been estimated from knowledge of the history of the breed, or from variation in the blood groups present. However, molecular genetic tools can be used to measure the diversity or *genetic distance* between breeds more objectively, to ensure that the most diverse breeds are conserved, and that the conserved population is based on individuals which are as dissimilar as possible. These techniques can also be used to target breeds or individuals carrying particularly rare or important alleles, where these are known (see, for example, FAO, 2015).

Genes conferring resistance to viruses, resistance to insects and herbicide tolerance have been transferred into plants from viruses, bacteria and other plant species. This supports the need for

wider conservation efforts across phyla and species, not just those of immediate concern. While these techniques have yet to be applied widely to animals for agricultural purposes, this remains a future possibility. However, this approach may be supplanted by *gene editing* techniques, outlined in Chapter 5, that allow targeted changes in base sequence within the breed and species concerned.

Methods of conservation

The main methods of conservation which have been proposed or used are:

- **Maintaining breeds in a commercial farm environment.** The advantages of this approach are that the genetic resources are still being utilized, they can be seen and enjoyed, the performance characteristics can be properly recorded, and the breeds have the opportunity to evolve, e.g. to develop resistance to new disease challenges, or to adapt to changes in husbandry. The disadvantages are that selection and genetic drift (the chance change in gene frequency which occurs across generations due to small population size) may result in unfavourable genetic changes, there is a risk of increasing inbreeding and hence homozygosity which is associated with reduced fitness, and the animals are at risk from disease or other natural disasters. Also, they are likely to be less productive, and so more costly to maintain than more popular breeds. This may increase the risk of individual breeders ceasing to keep rare breeds. Grants to farmers keeping rare breeds are used in some countries as a way of promoting this type of conservation.
- **Maintaining breeds in farm parks or other collections.** Many of the pros and cons are similar to the method above. However, there is potentially more control over the management of the population. Also, there may be greater opportunities for education, and to recoup some of the costs of conservation, by promoting public access. This approach, pioneered by the RBST, has become very popular in several countries.
- **Creating a gene pool.** This involves crossing several rare breeds together, then breeding them to maintain genetic variability (Maijala, 1970). It is probably an effective way of conserving genetic variation from two or three populations, but there is a greater risk of losing useful genes when more populations are combined. Although useful genes may be conserved with this approach, the 'identity' of different breeds is lost.
- **Frozen storage (*cryopreservation*) of semen, embryos, oocytes, somatic cells, *primordial germ cells* or DNA from breeds at risk.** These methods have the advantage that, after the initial investment, they are relatively inexpensive. Also, the population being conserved is free from unintended genetic change. During storage, frozen genetic material is at less risk from disease and natural disasters than live animals, but obviously at risk from technological failures. These risks can be minimized by splitting the material collected from a particular population and storing it in different locations. The disadvantages are that reproductive technologies are not uniformly successful for all individuals or all species, and the expertise is not always available in the places where it is needed most for collection or reinstatement. Also, cryopreservation prevents adaptation to changes in the production environment and to new disease challenges. If frozen semen is used as the only method of conservation, then several generations of backcrossing are

needed to reinstate the breed concerned. In contrast, breeds can be reinstated rapidly from frozen embryos. At the time of writing, techniques for recreating populations from somatic (body) cells, primordial germ cells (cells that give rise to eggs) or DNA are either still at an experimental stage or yet to be tested. For instance, scientists at the Roslin Institute, University of Edinburgh, have demonstrated experimentally an approach for regenerating rare breeds of chicken. This involves collecting and freezing primordial germ cells from rare breeds and transplanting these into eggs from hens genetically modified to be infertile, giving rise to chicks with DNA entirely from the rare breed. This so-called ‘frozen avairy’ provides a potentially efficient method of conserving avian genetic resources (Liu *et al.*, 2017).

- ***In silico* conservation.** Sequencing the whole genomes of breeds at risk provides a form of *in silico* conservation, potentially allowing reconstruction at a future date, by means yet to be tested.

When the costs of different current methods of conservation are compared, as well as their effectiveness in avoiding inbreeding, the combined use of live animals and frozen semen or embryos appears to be the best strategy (Lömker and Simon, 1994; FAO, 2019). However, integrating the use of some of the newer techniques is likely to be cost effective in the medium term. When conservation is based on live animal populations, or a combination of live animals and frozen genetic material, there are useful guidelines which help to maintain genetic variation in the population (de Rochambeau and Chevalet, 1990; Frankham, 1994; FAO, 1998, 2015, 2019). In summary, these are:

- Start with as variable a population as possible.
- Start with as large a population as possible.
- Turn over generations as slowly as possible.
- Use enough parents (especially males) to keep inbreeding at acceptable levels (what acceptable levels of inbreeding are, the effect of number of parents on rates of inbreeding, and mating designs to minimize inbreeding are discussed in more detail in Chapter 4).
- Minimize the variation in family sizes to reduce inbreeding – ideally each male parent should be replaced by a son, and each female by a daughter.
- Subdivide the breeding population to reduce inbreeding and genetic drift. In small populations it may be most effective to practice a form of *within-family selection*, i.e. to organize the population into ‘families’ in a notional circle, and keep female replacements in the family in which they were born, but move replacement males from the family in which they were born, to the next family in the circle, in the same sequence each year (see Fig. 3.16). In slightly larger populations it is more effective to mate males from one group to females of different groups in successive years.

Sometimes, unselected ‘control’ populations of animals are maintained as a benchmark against which to measure response to selection in an experimental or industry breeding scheme. The guidelines above are also helpful in establishing and maintaining such a population.

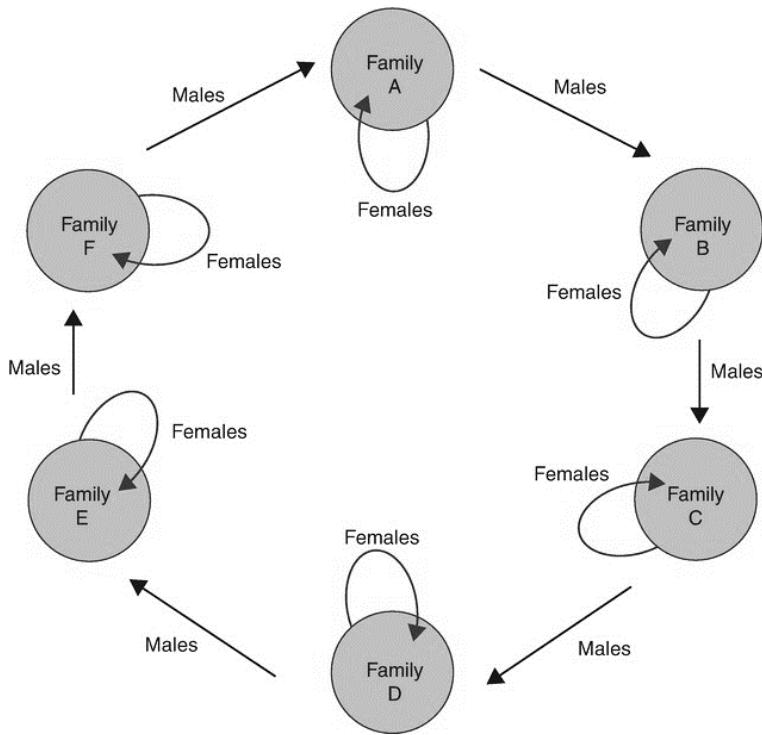


Fig. 3.16. A diagram illustrating the use of within-family selection to limit inbreeding. Female replacements remain in the family in which they were born, while males born in one family move to the next for mating. (This method was used to limit inbreeding in the unselected 'control' line in an SAC selection experiment using Suffolk sheep. The control line was maintained to allow estimation of response to selection by comparing the performance of animals in the two lines each year. The control line comprised 75 breeding females divided into 6 families, and 6 males were used per annum. After 10 years the level of inbreeding in this flock was similar to that in the selection line which was over double the size of the control line but did not use within-family selection.)

Summary

- Traditionally, three main strategies have been used for the genetic improvement of livestock. These are (i) selection between breeds or breed types; (ii) selection within breeds or breed types; and (iii) crossbreeding. New strategies, such as the ability to transfer genes within or between species, and gene editing, are becoming available as a result of developments in molecular genetics and reproductive biology.
- For any genetic improvement strategy to be effective, it is important to have a clear view of what the economically-important traits are. Then it is logical to choose the most appropriate breed or cross, based on their performance in these traits. It is then sensible to consider whether this pure breed, or component breeds of the cross, can be improved further by within-breed selection.
- The structure of livestock breeding industries in most industrialized nations is often described schematically as a pyramid with *elite* or *nucleus* breeders at the top, one or more middle tiers of purebred or crossbred *multipliers*, and a final tier of *commercial herds* or *flocks*, or end users. Ideally, market signals from commercial herds or flocks influence breeding decisions in the tiers above.
- Selection between breeds or strains can achieve dramatic and rapid genetic change when there are large genetic differences between populations in characteristics of economic or other importance. It is costly to replace whole flocks or herds of breeding females at once. In practice, changes are often made more gradually by grading up to the new breed. It is important that

choices among breeds are made on the basis of well-designed objective comparisons of performance in the relevant production environment.

- Selection within breeds involves comparing animals of the same breed and mating the preferred animals to produce the next generation. This process is usually repeated each generation, and as long as there is genetic variation in the characters under selection, this produces changes in each generation.
- Objective selection within breeds or strains is intended to increase the average level of additive genetic merit or breeding value of the population. Ideally, the steps involved are: (i) deciding on the breeding goal; (ii) deciding on the selection criterion; (iii) designing the breeding programme, e.g. numbers of males and females selected annually, ages at mating; (iv) implementing the programme, i.e. doing the routine recording, evaluation and mating of animals; and (v) monitoring progress and redesigning the programme, as necessary.
- Objective selection depends on having records of performance on the candidates for selection, or their relatives, or both. In most industrialized countries, the majority of cattle and sheep breeders use recording schemes which are operated by regional or national agencies who specialize in recording and evaluation. Although these schemes are sometimes less flexible and more expensive than ‘home-grown’ schemes, they encourage discipline and consistency in recording, they often allow access to more sophisticated methods of evaluation, and they can improve communication with customers.
- Annual rates of genetic improvement in polygenic traits depend on the *selection differential*, the *heritability* of the trait under selection, and the *generation interval*. Generally speaking, the higher the selection differential and heritability, and the lower the generation interval, the higher the annual rate of genetic improvement.
- Profitability usually depends on several different animal characteristics rather than any single trait. Hence, in genetic improvement programmes animals are usually selected on a combination of traits. This can be achieved by tandem selection, the use of independent culling levels or, most efficiently, by index selection.
- Crossbreeding involves mating animals of different breeds, lines or species. It is usually used for one or more of the following reasons: (i) to improve the overall efficiency of a production system by crossing breeds which each have high genetic merit in different traits (complementarity); (ii) to produce individual animals of intermediate performance between that of two more extreme parent breeds; (iii) for grading up to a new breed or strain; (iv) as an intermediate step in the creation of a new synthetic or composite breed; (v) to introduce new variation to numerically small breeds; (vi) to introduce a single gene for a favourable characteristic to an existing breed (introgression); or (vii) to exploit heterosis or hybrid vigour.
- Heterosis is defined as the advantage in performance above the mid-parent mean. It is most useful when it leads to the average performance of crossbred animals exceeding that of the best parent breed. Individual heterosis directly influences the performance of crossbred animals themselves. Maternal and paternal heterosis arise when dams or sires are crossbred, and the effects are often measured in terms of improved reproductive efficiency or improved performance of offspring.
- Different systems of crossing lead to different proportions of individual, maternal and paternal heterosis being maintained. However, the most appropriate system of crossing depends not only

on this, but also on the additive merit of the breeds available, and the absolute level of performance of crossbreds.

- Each of the strategies for genetic improvement depends on genetic variation, so it is important for the success of future improvement programmes that genetic variation is used in a sustainable way. However, there are aesthetic as well as biological reasons for conservation of genetic resources.
- It is difficult to decide which breeds should be conserved. International organizations, such as the FAO, are attempting to identify those breeds most in need of conservation programmes and provide guidelines to implement these. Molecular genetic tools may help to ensure that the most diverse breeds and individuals are conserved.
- The main methods of conservation which have been proposed or used are: (i) maintaining breeds in their normal farm environment; (ii) maintaining breeds in farm parks or other collections; (iii) creating a gene pool; and (iv) frozen storage of semen, embryos, oocytes, somatic or primordial germ cells or DNA from rare breeds. When the costs of these different methods of conservation are compared, as well as their effectiveness in avoiding inbreeding, the combined use of live animals and frozen semen or embryos is usually the best current strategy.
- When conservation is based on live animal populations, it helps in maintaining genetic variation to: (i) start with as variable a population as possible; (ii) start with as large a population as possible; (iii) turn over generations as slowly as possible; (iv) use enough parents (especially males) to keep inbreeding at acceptable levels; (v) minimize the variation in family sizes to reduce inbreeding; and (vi) subdivide the breeding population to reduce inbreeding and genetic drift.

References

- ACGG (2019) African Chicken Genetic Gains project. Available at: <https://africacgg.net/> (accessed 17 May 2019).
- ADGG (2019) African Dairy Genetic Gains project. Available at: <https://africadgg.wordpress.com/> (accessed 17 May 2019).
- Beatson, P.R. (1993) Coopworths: making a concept happen. In: Blair, H.T. and McCutcheon, S.N. (eds) *Proceedings of the A.L. Rae Symposium on Animal Breeding and Genetics*, Massey University, Palmerston North, New Zealand, pp. 154–159.
- Bunning, H., Wall, E., Chagunda, M.G.G., Banos, G. and Simm, G. (2018) Heterosis in cattle crossbreeding schemes in tropical regions: Meta-analysis of effects of breed combination, trait type and climate on level of heterosis. *Journal of Animal Science* 97, 29-34. DOI: 10.1093/jas/sky407.
- Cameron, N.D., Smith, C. and Deeble, F.K. (1983) Comparative performance of crossbred ewes from three crossing sire breeds. *Animal Production* 37, 415-421. DOI: 10.1017/S0003356100002038.
- Cartwright, T.C. (1982) Objectives in beef cattle improvement. In: Barton, R.A. and Smith, W.C. (eds) *Proceedings of the World Congress on Sheep and Beef Cattle Breeding*, Vol. II. Dunmore Press, New Zealand, pp. 19–27.
- Cassady, J.P., Young, L.D. and Leymaster, K.A. (2002) Heterosis and recombination effects on pig growth and carcass traits. *Journal of Animal Science* 80, 2286–2302.
- CBD (2019) The Nagoya Protocol on Access and Benefit-sharing. Available at: <https://www.cbd.int/abs/> (accessed 19 May 2019).
- Cundiff, L.V. and Gregory, K.E. (1977) *Beef Cattle Breeding*. United States Department of Agriculture, Agriculture Information Bulletin No. 286. USDA, Washington, DC.
- Cundiff, L.V., Gregory, K.E. and Koch, R.M. (1982) Effects of heterosis in Hereford, Angus and Shorthorn rotational crosses. In: *Beef Research Program Progress Report No. 1.* (ARM-NC-21) pp. 3–5. Roman L. Hruska U.S. Meat Animal Research Center, Nebraska.
- Cundiff, L.V., Gregory, K.E., Koch, R.M. and Dickerson, G.E. (1986) Genetic diversity among cattle breeds and its use to increase beef production efficiency in a temperate environment. *Proceedings of the 3rd World Congress on*

- Genetics Applied to Livestock Production* Vol. IX, pp. 271–282. Available at: <http://www.wcgalp.org/proceedings/1986> (accessed 25 June 2020).
- Cunningham, E.P. and Syrstad, O. (1987) *Crossbreeding Bos indicus and Bos taurus for Milk Production in the Tropics*. FAO Animal Production and Health Paper No. 68, Food and Agriculture Organization of the United Nations, Rome. Available at: <http://www.fao.org/3/t0095e/T0095E00.htm> (accessed 28 August 2020).
- Davis, G.P. and Arthur, P.F. (1994) Crossbreeding large ruminants in the tropics: current knowledge and future directions. *Proceedings of the 5th World Congress on Genetics Applied to Livestock Production*, Vol. 20, pp. 332–339. Available at: <http://www.wcgalp.org/proceedings/1994> (accessed 25 June 2020).
- Defra (2019) Farm Animal Genetic Resources Committee. Available at: <https://www.gov.uk/government/groups/farm-animal-genetic-resources-committee-fangr> (accessed 10 January 2019).
- de Rochambeau, H. and Chevalet, C. (1990) Genetic principles of conservation. *Proceedings of the 4th World Congress on Genetics Applied to Livestock Production*, Vol. XIV, pp. 434–442. Available at: <http://www.wcgalp.org/proceedings/1990> (accessed 25 June 2020).
- Dickerson, G.E. (1973) Inbreeding and heterosis in animals. *Journal of Animal Science* 1973, 54–77. DOI : 10.1093/ansci/1973.Symposium.54.
- Dickerson, G.E. (1974) Evaluation and utilization of breed differences. *Proceedings of Working Symposium on Breed Evaluation and Crossing Experiments in Farm Animals*. Instituut voor Veeteeltkundig Onderzoek, 'Schoonoord', Zeist, Netherlands, pp. 7–23.
- Dickerson, G.E. (1978). Animal size and efficiency: basic concepts. *Animal Production* 27, 367–379. DOI: 10.1017/S0003356100036278.
- Falconer, D.S. and Mackay, T.F.C. (1996) *Introduction to Quantitative Genetics*, 4th edn. Longman, Harlow, UK.
- FAO (1998) Secondary Guidelines for Development of Farm Animal Genetic Resources Management Plans: Management of small populations at risk. Available at: <http://www.fao.org/3/a-w9361e.pdf> (accessed 08 January 2019). Food and Agriculture Organization of the United Nations, Rome.
- FAO (2007) The State of the World's Animal Genetic Resources for Food and Agriculture. Rischkowsky, B. and Pilling, D. (eds). FAO, Rome. Available at: <http://www.fao.org/3/a-a1260e.pdf> (accessed 26 October 2019).
- FAO (2015) The Second Report on the State of the World's Animal Genetic Resources for Food and Agriculture. Food and Agriculture Organization of the United Nations, Rome. Available at: <http://www.fao.org/3/a-i4787e.pdf> (accessed 10 January 2019).
- FAO (2019) Genetic Resources. Food and Agriculture Organization of the United Nations, Rome. Available at: <http://www.fao.org/genetic-resources/en/> (accessed 10 January 2019).
- Fell, H.R. (1979) *Intensive Sheep Management*. Farming Press, Ipswich, UK.
- Frankham, R. (1994) Conservation of genetic diversity for animal improvement. *Proceedings of the 5th World Congress on Genetics Applied to Livestock Production*, Vol. 21, pp. 385–392. Available at: <http://www.wcgalp.org/proceedings/1994> (accessed 28 August 2020).
- Gjedrem, T. and Rye, M. (2018) Selection response in fish and shellfish: a review. *Reviews in Aquaculture* 10, 168–179. DOI: 10.1111/raq.12154.
- Gregory, K.E., Cundiff, L.V. and Koch, R.M. (1982) Comparison in crossbreeding systems and breeding stocks used in suckling herds of continental and temperate areas. *Proceedings of the 2nd World Congress on Genetics Applied to Livestock Production*, Vol. V. pp. 482–503. Available at: <http://www.wcgalp.org/proceedings/1982> (accessed 28 August 2020).
- Gregory, K., Cundiff, L.V. and Koch, R.M. (1993) Composite breeds – what does the research tell us? *Range Beef Cow Symposium* 207. Available at: <http://digitalcommons.unl.edu/rangebeefcowsymp/207> (accessed 17 May 2019).
- Harris, B.L., Clark, J.M. and Jackson, R.G. (1996) Across breed evaluation of dairy cattle. *Proceedings of the New Zealand Society of Animal Production* 56, pp. 12–15. Available at: <http://www.nzsap.org/system/files/proceedings/1996/ab96005.pdf> (accessed 10 January 2019).
- Harris, D.L., Stewart, T.S. and Arboleda, C.R. (1984) *Animal Breeding Programs: A Systematic Approach to their Design*. Advances in Agricultural Technology, Publication No. AAT-NC-8, Agricultural Research Service, North Central Region, US Department of Agriculture, Peoria, Illinois.
- Innovis (2019) Innovis – Breeding Sheep. Available at: www.innovis.org.uk (accessed 1 August 2019).
- Kelso (2019) Kelso. Available at: <http://kelso.co.nz/> (accessed 1 August 2019).
- Kempster, A.J., Cook, G.L. and Southgate, J.R. (1982) A comparison of different breeds and crosses from the suckler herd. 2. Carcass characteristics. *Animal Production* 35, 99–111. DOI: 10.1017/S0003356100000866.
- Kempster, A.J., Croston, D., Guy, D.R. and Jones, D.W. (1987) Growth and carcass characteristics of crossbred lambs by ten sire breeds, compared at the same estimated subcutaneous fat proportion. *Animal Production* 44, 83–98. DOI: 10.1017/S0003356100028099.
- Leachman (2019) Leachman Stabilisers. Available at: <https://www.leachman.com/leachman-stabilizers> (accessed 1 August 2019).
- Liu, C., Chang, I.-K., Khazanehdari, K.A., Thomas, S., Varghese, P. et al. (2017) Uniparental chicken offsprings derived from oogenesis of chicken primordial germ cells (ZZ). *Biology of Reproduction* 96, 686–693. DOI: 10.1093/biolreprod.116.144253.

- Livestock Improvement, (1996) *Dairy Statistics 1995-1996*. Livestock Improvement Corporation Limited, New Zealand Dairy Board, Hamilton, New Zealand.
- Livestock Improvement Corporation Ltd, and Dairy NZ Ltd, (2018). *New Zealand Dairy Statistics 2017-18*. LIC, Private Bag 3016, Hamilton 3240, New Zealand; Dairy NZ, Private Bag 3221, Hamilton 3240, New Zealand.
- Lömker, R. and Simon, D.L. (1994) Costs of and inbreeding in conservation strategies for endangered breeds of cattle. *Proceedings of the 5th World Congress on Genetics Applied to Livestock Production*, Vol. 21, pp. 393–396. Available at: <http://www.wcgalp.org/proceedings/1994> (accessed 23 August 2020).
- Lush, J.L. (1945) *Animal Breeding Plans*, 3rd edn. Iowa State College Press, Ames, Iowa.
- Maijala, K. (1970) Need and methods of gene conservation in animal breeding. *Annales de Génétique et de Sélection Animale* 2, 403–415. DOI: 10.1186/1297-9686-2-4-403.
- Marshall, K. (2014) Optimizing the use of breed types in developing country livestock production systems: A neglected research area. *Journal of Animal Breeding and Genetics* 131, 329–340. DOI: 10.1111/jbg.12080.
- Marshall, K., Tebug, S., Juga, J., Tapio, M. and Missohou, A. (2016) Better Dairy Cattle Breeds and Better Management Can Improve the Livelihoods of the Rural Poor in Senegal. ILRI Research Brief 65. ILRI, Nairobi. <https://cgspace.cgiar.org/handle/10568/72865>.
- Marshall, K., Salmon, G.R., Tebug, S., Juga, J., MacLeod, M. et al. (2020) Net benefits of smallholder dairy cattle farms in Senegal can be significantly increased through the use of better dairy cattle breeds and improved management practices. *Journal of Dairy Science* 103, 8197–8217.
- Mayala, K. (1974) Breed evaluation and crossbreeding in sheep. A summarising report. *Proceedings of Working Symposium on Breed Evaluation and Crossing Experiments in Farm Animals*. Instituut voor Veeteeltkundig Onderzoek, 'Schoonoord', Zeist, Netherlands, pp. 389–405.
- Meat and Livestock Commission. (1988) *Sheep in Britain*. MLC, Milton Keynes, UK.
- Mueller, J., Rischkowsky, B., Haile, A., Philipsson, J., Mwai, O. et al. (2015) Community-based livestock breeding programmes: essentials and examples. *Journal of Animal Breeding and Genetics* 132, 155–168. DOI: 10.1111/jbg.12136.
- Nicholas, F.W. (1987) *Veterinary Genetics*. Oxford University Press, Oxford.
- Parker, A.G.H. and Rae, A.L. (1982) Underlying principles of co-operative group breeding schemes. In: Barton, R.A. and Smith, W.C. (eds) *Proceedings of the World Congress on Sheep and Beef Cattle Breeding*, Vol. II. Dunmore Press, New Zealand, pp. 95–101.
- Simm, G. (1998). *Genetic Improvement of Cattle and Sheep*. CAB International, Wallingford, UK.
- Simm, G., Conington, J. and Bishop, S.C. (1994) Opportunities for genetic improvement of sheep and cattle in the hills and uplands. In: Lawrence, T.L.J., Parker, D.D. and Rowlinson, P. (eds) *Livestock Production and Land Use in Hills and Uplands*. Occasional Publication No. 18, British Society of Animal Production, pp. 51–66. DOI: 10.1017/S0263967X00001506.
- Smith, C. (1964) The use of specialised sire and dam lines in selection for meat production. *Animal Production* 6, 337–344. DOI: 10.1017/S0003356100022133.
- Smith, C. (1984) Rates of genetic change in farm livestock. *Research and Development in Agriculture* 1, 79–85.
- Snedecor, G.W. and Cochran, W.G. (1989) *Statistical Methods*, 8th edn. Iowa State University Press, Ames, Iowa.
- Southgate, J.R., Cook, G.L. and Kempster, A.J. (1982) A comparison of different breeds and crosses from the suckler herd. 1. Live-weight growth and efficiency of food utilization. *Animal Production* 35, 87–98. DOI: 10.1017/S0003356100000854.
- Thiessen, R.B., Hnizdo, E., Maxwell, D.A.G., Gibson, D. and Taylor, C.S. (1984) Multibreed comparisons of British cattle. Variation in body weight, growth rate and food intake. *Animal Production* 38, 323–340. DOI: 10.1017/S0003356100041520.
- Timon, V.M. (1974) The evaluation of sheep breeds and breeding strategies. *Proceedings of Working Symposium on Breed Evaluation and Crossing Experiments in Farm Animals*. Instituut voor Veeteeltkundig Onderzoek, 'Schoonoord', Zeist, Netherlands, pp. 367–387.
- Wainwright, W., Ahmadi, B., McVittie, A., Simm, G. and Moran, D. (2019) Prioritising support for cost effective rare breed conservation using multi-criteria decision analysis. *Frontiers in Ecology and Evolution* 7, 110. DOI: 10.3389/fevo.2019.00110.
- Wiener, G. (1994) *Animal Breeding*. The Tropical Agriculturalist series. Macmillan, London and CTA, Wageningen, Netherlands.
- Williams, S.M., Price, S.E. and Siegel, , P.B., (2002) Heterosis of growth and reproductive traits in fowl. *Poultry Science* 81, 1109–1112. DOI: 10.1093/ps/81.8.1109.

3.1.2

Simm et al. (2020)

Page 64-80

4. What Affects Response to Selection Within Breeds?

Introduction

Over the last few decades, in higher-income countries, objective selection within breeds has been practised widely in pig and poultry breeding, and to a slightly lesser extent in dairy cattle and finfish breeding. It has been used less widely still in beef cattle, sheep and goat breeding. There are several reasons for this. These include the ‘biological advantages’ that allow faster rates of genetic improvement in pigs and poultry, as outlined in Chapter 3, and so higher rates of return on investment. Pig and poultry breeding in higher-income countries has become concentrated in a small number of global companies, which sell breeding stock into a large and very competitive sector that sends clear market signals. So, setting breeding goals has been more straightforward. In contrast, market signals have often been less clear in beef cattle and sheep production, and in some countries commercial breeding objectives have been obscured by fashions in the show ring, until recently. Those responsible for breeding decisions in pig and poultry breeding companies have been committed to objective methods, and less influenced by traditional breeding practices than cattle and sheep breeders. In contrast, beef and sheep breeding is often in the hands of individual breeders, many of whom have a long family history in the profession. Some have been sceptical about new approaches, which is understandable if traditional methods have served them well from a financial perspective. However, this is changing, with much wider recognition of the value of objective methods. This tradition was particularly strong in Britain, where so many breeds of livestock were developed and exported world-wide – though history shows the dangers of ‘resting on your laurels’ in livestock breeding. Many breeds which were very popular even a few decades ago are minority breeds today.

Dairy cattle breeding has fallen between these two extremes. Perhaps this is because market signals have been clearer than in beef cattle and sheep production, and because breeders, and their customers, are confronted each milking time with visible evidence of the link between their breeding decisions and profitability. For commercial beef and sheep producers, selection between breeds and crossbreeding have been easier strategies for improvement, because there are often more obvious differences between breeds and crosses than between individual bulls or rams of the same breed at a sale. Also, there are often easily identified ‘markers’, like coat colour, for particular breeds or crosses of proven merit.

Objective breeding programmes are a newer concept in finfish farming, and especially in shellfish farming. However, they have been adopted rapidly in some species (e.g. salmon) and are being adopted in others, benefiting from experience in terrestrial species. In several species, especially salmon, there is a growing trend towards a small number of broodstock producers operating across major production countries, as in pigs and poultry. In some cases, the same companies are involved in multiple species.

It would be too simplistic to set up pig and poultry breeding as a model to which other breeders should aspire, without qualification. There is no doubt that pig and poultry breeders have employed objective breeding techniques very successfully to improve major performance traits. However, at least in their earlier years, it is fair to say that most had rather narrow breeding goals. In some cases, this contributed to problems of ‘functional fitness’, such as leg weakness in pigs and poultry, and the inability of the males of some strains of turkey to mate naturally because of excessive breast muscle development. There is now a growing awareness that there are animal welfare, ethical and economic arguments against the pursuit of very narrow breeding goals. While this increases the importance of choosing appropriate breeding goals, and paying attention to the functional fitness of selected animals, it should not detract from the value of objective selection. The success of objective selection methods in pigs and poultry is sometimes attributed to the intensive systems in which they are often kept. Although performance recording may be easier in these systems, intensive production systems are not a prerequisite for successful selection programmes. If cattle and sheep breeders are going to meet the more exacting demands of their local customers, and if animal agriculture is going to help to meet the demands of a growing global population, then it is vital that there is wider use of relevant objective methods of within-breed selection. The key to this will be a better understanding of the procedures involved, their value and their limitations.

Hence, the purpose of this and the next three chapters is to expand on the introduction to within-breed improvement given in Chapter 3. This chapter deals with predicting response to selection and designing breeding programmes to achieve the maximum response to selection. Chapter 5 examines tools and technologies used in breeding, while Chapter 6 covers analysing genetic variation. Chapter 7 deals with predicting the breeding values of individual animals – a vital component in achieving responses once the overall design of a breeding scheme has been decided.

Factors Affecting Rates of Genetic Gain

The factors affecting rates of genetic gain were introduced in Chapter 3. These are easiest to understand if we consider the special case where animals are selected on a single measurement of their own performance in one trait, ignoring information from relatives. This is often called individual or mass selection. It is also simplest to consider the process of selection in a single closed flock or herd – that is, a flock or herd which is producing all of its own replacement breeding stock, rather than buying these in from other flocks or herds. Each year new animals are born, their performance is recorded, and the best performing animals are selected and retained. Once they reach normal breeding age, they replace those adult males or females in the breeding herd or flock which have died of natural causes, or which the breeder wishes to cull and replace by superior animals. Although this is an annual process, it is easiest to follow if we concentrate initially on a single group of recorded animals and their offspring.

For example, let us consider a group of lambs of both sexes, each with a record of 20-week weight (generation 1). From this group, the lambs of each sex with the highest 20-week weight are selected for breeding to each other. The same performance trait, 20-week weight, is recorded eventually in their offspring (generation 2). The response is measured by comparing the average performance of all the lambs in generation 1 with the average performance of all of the lambs in generation 2 (i.e.

the offspring of those animals selected for breeding from generation 1). In this case, the response to selection (R) per generation depends on just two things: the selection differential achieved (S), and the heritability of the trait selected on (h^2), as we saw in Chapter 3 (after Lush, 1945):

$$R = S \times h^2$$

Selection differential

The selection differential is the difference between the mean performance of selected animals (i.e. those identified to become parents of the next generation) and the overall mean of the group of animals from which they were selected. To continue the example above, it is the difference between the mean performance of selected animals in generation 1, and the overall mean of generation 1 animals. This is illustrated in Fig. 4.1. Because there are usually unequal numbers of male and female parents selected, but each sex eventually contributes half of the genes to the offspring, it is important to calculate selection differentials separately for males and females, and then average them to get an overall selection differential. In this example, the selection differential achieved in males is +10 kg, while that achieved in females is +2 kg, so the average selection differential achieved is +6 kg ($[10 + 2] / 2$). Figure 4.1 illustrates that the fewer animals selected for high performance, the further away from the mean they will be, and so the higher the selection differential will be. A further point to note is that the wider the variation in performance of the group, the higher the selection differential which can be achieved. (Although, as we will see later, this is only of value if there is a lot of genetic, rather than environmental, variation).

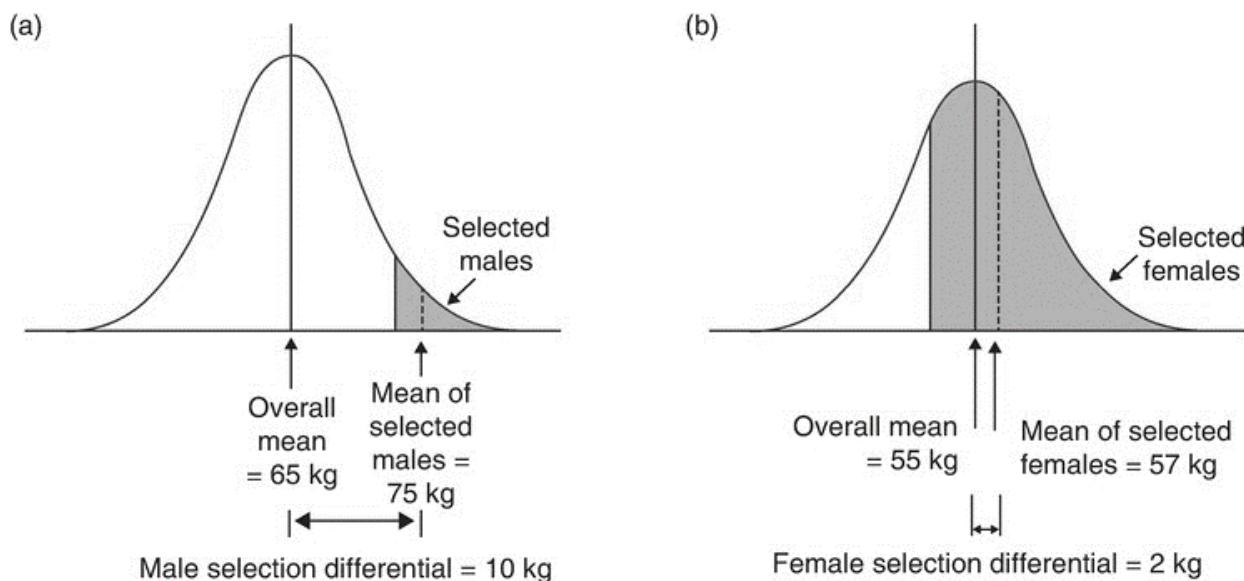


Fig. 4.1. (a) The selection differential achieved in males by selecting the heaviest ram lambs from a performance-recorded group at 20 weeks of age. The selection differential among males is the average weight of selected males minus the average of the group of males as a whole. (b) The selection differential achieved by selecting a larger proportion of ewe lambs from a recorded group, at the same age.

Heritability

The second factor affecting the response to selection per generation is the heritability of the trait concerned. There are several ways of defining heritability. As mentioned in the last chapter, the simplest definition is that the heritability is the proportion of superiority of parents in a trait (i.e. the proportion of the selection differential) which, on average, is passed on to offspring. So, if the heritability of a trait is high, a lot of the superiority of the parents is a result of the genes they carry and we can expect much of this superiority to be passed on to offspring. If the heritability of a trait is low, only a little of the superiority of parents is a result of the genes they carry, and so only a little of this superiority will get passed on to offspring. Heritabilities are expressed as proportions from 0 to 1, or as percentages from 0 to 100%.

It will be easiest to understand selection differentials and heritabilities, and how they act together to influence response, by returning to the sheep example. In the case above, the average selection differential achieved across both sexes was +6 kg. If the heritability of live weight at 20 weeks of age

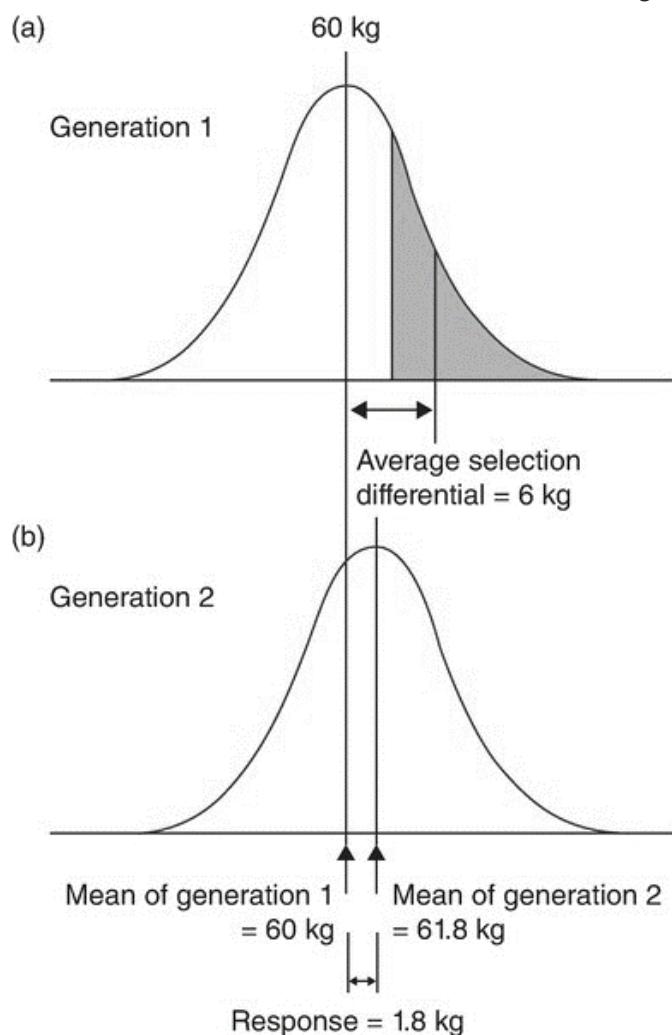


Fig. 4.2. Expected response to selection on 20-week weight in sheep, with an average selection differential of +6 kg, and a heritability of 0.3.

is 0.3 or 30%, then we expect 30% of this 6 kg superiority to be passed on to offspring, on average. So, in this example we expect a 1.8 kg improvement in 20-week weight of the next generation ($0.3 \times 6 \text{ kg}$). This is illustrated in Fig. 4.2. It is important to note that selection achieves an increase in the average performance of the next generation – it does *not* add 1.8 kg to the weight of every lamb.

So how do we know what the heritability of a particular trait is in the first place? It may sound circular, but the simplest way of estimating the heritability for a particular trait is to select parents on this trait, measure the selection differential achieved, mate them, measure the same trait in their offspring and then calculate, using the regression techniques outlined in Chapter 2, how much of the superiority of parents was passed on to offspring. There are other more sophisticated methods which involve measuring the degree of resemblance in performance between different classes of relatives; these are discussed in Chapter 6.

Also, there is an alternative definition of heritability, which follows from the discussion of different types of variation in Chapter 2. The heritability of a trait is the

additive genetic variation in that trait (or the variation in breeding values), expressed as a proportion of the total phenotypic variation in that trait. The fact that the heritability is the ratio of two variances explains why it is abbreviated to h^2 . The square root of the heritability, abbreviated to h , is the ratio of the *additive genetic standard deviation* of a trait to the phenotypic standard deviation. Using the abbreviations introduced in Chapter 2 (after Lush, 1945):

$$\text{Heritability} = \text{VAR}_A / \text{VAR}_P$$

Distributions of VAR_A and VAR_P for two traits with different heritabilities are shown in Fig. 4.3. The inner distributions show the proportion of the total phenotypic variation which is a result of additive genetic variation, or variation in breeding values. In one case, additive genetic variation accounts for 10% of the total phenotypic variation so, as we have seen, the heritability is 10%. In the other case a higher proportion of the total variation, 50%, is additive genetic variation, so the heritability is 50%. The graphs illustrate that when the heritability of a trait is low, the distribution of breeding values will be narrower, relative to phenotypic variation, than when the heritability of the trait is high. Table 4.1 shows typical examples of heritabilities for some traits of interest in cattle and sheep.

The ability to predict response to selection is important in planning breeding programmes, and especially in comparing alternative schemes. The problem with the approach described so far is that it predicts response per generation, but generations can be of variable length. For instance, in the sheep example, if the selected males and females had been mated at 7 months of age, the earliest age possible in most breeds, then their offspring would be born when they were about a year old. If, on the other hand, the selected animals were mated at 19 months of age, the offspring would be born a year later, when the parents were 2 years old. So, two breeders may be achieving equal selection differentials and equal responses per generation, but achieving these responses in quite different lengths of time. The solution to this problem is to express predicted responses *per annum* rather than per generation. To do this, we need to know the *generation interval* in the flock or herd concerned.

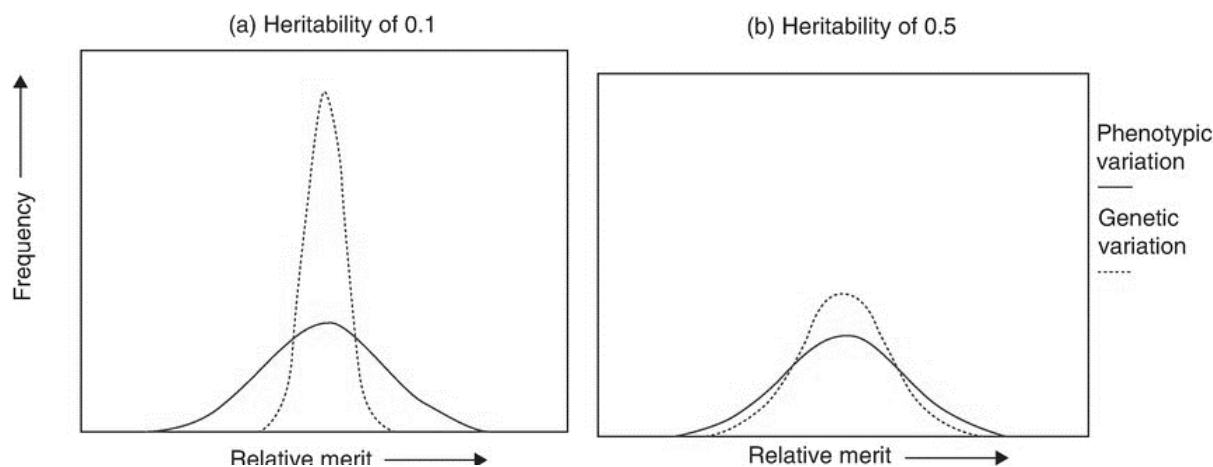


Fig. 4.3. Distributions of additive genetic variation (VAR_A) and phenotypic variation (VAR_P) for two traits with different proportions of additive genetic variation, and hence different heritabilities. (a) Heritability of 0.1 or 10%. (b) Heritability of 0.5 or 50%.

Table 4.1. Typical heritabilities for some traits of economic importance. See Chapters 8 to 13 for average values for these and many other traits from comprehensive literature searches

Species/breed	Trait	Typical heritability values
Dairy cattle	Milk yield	25–40%
	Milk fat %	50–60%
	Milk protein %	50–60%
	Various measures of fertility	1–5%
	Incidence of various production diseases	0–25%
Beef cattle	Birth weight	25–35%
	Weaning weight	20–30%
	Yearling weight	30–40%
	Carcass fat depth	35–50%
	Calving ease	5–15%
Sheep	Birth weight	10–25%
	Weaning weight	15–25%
	Yearling weight	20–35%
	Carcass fat depth	25–35%
	Various combined measures of fertility and lamb output	5–15%
	Fleece weight	30–40%
	Fibre diameter	45–55%
Chickens	Growth rate	30–60%
	Feed conversion ratio	15–30%
	Resistance to range of diseases	5–50%
	Egg number	15–60%
	Egg weight	30–70%
Pigs	Litter size	5–10%
	Average daily gain	10–30%
	Backfat depth	40–65%
	Feed conversion ratio	15–30%
Atlantic salmon	Growth rate	40–60%
	Carcass characteristics	10–50%
	Resistance to range of diseases	10–60%

Generation interval

The generation interval (abbreviated L) is the average age of parents when their offspring are born. Because there are unequal numbers of male and female parents, and they often breed at different ages, generation intervals have to be calculated separately for each sex of parent, and then averaged. An additional complication is that in most flocks or herds there are different proportions of animals in each age group. Also, if the fertility of parents of different age groups differs, then this too must be taken into account. This means that we cannot simply average the parents' age classes, we must weight them according to the number of offspring produced by each age class. For example, if we had a flock of ewes comprising 50% 2-year olds and 50% 3-year olds, the average age would simply be 2.5 years. If, as is usually the case, we had a higher proportion of young animals in the flock to allow for natural wastage, then we could not simply average the age classes. If we had 60% 2-year olds and 40% 3-year olds, the average age of the ewes would be $(0.6 \times 2) + (0.4 \times 3) = 2.4$ years. If these two age groups produce equal numbers of offspring per ewe, then the female

generation interval would simply be 2.4 years. If the youngest ewes are less prolific than older ones, so that 55% of lambs come from two year old ewes, and 45% from three year old ewes, then the female generation interval would be 2.45 years ($[0.55 \times 2] + [0.45 \times 3] = 2.45$). A more realistic example is given in Table 4.2. As the generation interval can differ for males and females, but each contribute equally to the next generation, it is averaged over these.

Table 4.2. Calculating the average generation interval in a flock of sheep of mixed ages.

(a) *Calculating the female generation interval. In this example, the 100-ewe flock has a typical age distribution, with progressively fewer ewes in successive age groups, as a result of natural wastage. The average litter size is 1.6 lambs, but prolificacy is lowest in the youngest ewes, it is highest in ewes of intermediate age, and then declines slightly in the oldest ewes. Calculating the female generation interval involves recording the number of ewes in each age group at lambing (column 2), calculating the number of lambs (column 3) and then the proportion of lambs (column 4) from ewes in each age group. Each ewe age is then multiplied by the relevant proportion of lambs (column 1 \times column 4 = column 5), to get the contribution of this age group to the average weighted age. These values are then summed over all age groups to get a weighted average age, which is the female generation interval.*

Age of ewes at lambing (years)	Number of ewes in this age group	Number of lambs from ewes in this age group	Proportion of lambs from ewes in this age group (approx.)	Contribution of ewes in this age group to weighted average age (col. 1 \times col. 4)
2	25	35	0.22	0.44
3	22	37	0.23	0.69
4	20	34	0.21	0.84
5	18	30	0.19	0.95
6	15	24	0.15	0.90
Total	100	160	1.00	3.82
Weighted average age of ewes = 3.82 years (approx.)				
Female generation interval = 3.82 years (approx.)				

(b) *Calculating the male generation interval. In this example four rams are used in the flock, and they are mated to groups of ewes of mixed ages. Hence the two age groups of rams have roughly equal numbers of lambs born. Weighted average ages are calculated as described above for ewes.*

Age of rams at birth of their offspring (years)	Number of rams in this age group	Number of lambs from rams in this age group	Proportion of lambs from rams in this age group (approx.)	Contribution of rams in this age group to weighted average age (col. 1 \times col. 4)
1	2	78	0.49	0.49
2	2	82	0.51	1.02
Total	4	160	1.00	1.51
Weighted average age of rams = 1.51 years (approx.)				
Male generation interval = 1.51 years (approx.)				

(c) *Calculating the average generation interval.*

Female generation interval = 3.82 years (approx.)
Male generation interval = 1.51 years (approx.)
Average generation interval = 2.67 years (approx.)

To predict responses to selection *per annum*, rather than per generation, we simply divide the formula shown above for response per generation by the average generation interval (L):

$$R(\text{per annum}) = \frac{S \times h^2}{L}$$

So, if we selected on 20-week weight each year in a flock with the age structure shown in Table 4.2, and achieved the selection differentials shown in Fig. 4.1 each year, we would have:

$$S = 6 \text{ kg},$$

$$h^2 = 0.3, \text{ and}$$

$$L = 2.67 \text{ years}.$$

So, we would expect to get a response *per annum* of:

$$R = \frac{6 \times 0.3}{2.67}$$

$$= 0.67 \text{ kg per annum} \text{ (approximately).}$$

A more useful formula for predicting response

As mentioned already, it is very useful to be able to predict rates of response to selection, in order to compare alternative breeding programmes – for example, to choose the design likely to maximize annual response. At first sight it appears that we need to know quite a lot of information in advance in order to predict response. We need to know the heritability of the trait under selection, but for most traits there are already estimates available which we can use. We also need to know the generation interval, but usually this can be predicted accurately from the expected age distribution of parents, together with typical levels of reproductive performance for parents of each age group. A greater limitation of the formula used so far is that we need to know the selection differential in advance in order to predict response. It is easy to calculate a selection differential once animals have been performance recorded and parents have been selected, but how can we predict what selection differential is likely to be achieved before embarking on a breeding programme?

We saw in Chapter 2 that one of the features of normal distributions is that we can predict remarkably accurately what proportion of animals will fall in different ranges of performance around the mean. For example, about 68% of animals will have performance in the range from 1 standard deviation below the mean to 1 standard deviation above the mean, 95% of animals fall in the range -2 to +2 standard deviations, and so on. So, if we know how many standard deviations better than average our selected animals are, and we know the standard deviation of the trait of interest (i.e.

how many kg of live weight or litres of milk each standard deviation is ‘worth’), we can predict the selection differentials achieved.

If we know the total number of animals recorded, and the number selected from this group, there are standard statistical tables which allow us to express the merit of selected animals in standard deviation units above the mean. A table of these values is shown at the end of the book. Table 4.3 shows part of it. This table shows, for example, that if we select the best 5 rams out of a total of 75 that are performance recorded, on average they are expected to be 1.893 standard deviations above the mean of all 75 rams. If the standard deviation of 20-week weight is about 5 kg, then we expect the average weight of the five selected rams to be about 9.5 kg heavier than the average of all 75 recorded rams ($1.893 \times 5 \text{ kg} = 9.465 \text{ kg}$).

The superiority of animals selected, expressed in standard deviation units, is called the standardized selection differential or *selection intensity* (abbreviated i). Table 4.3 shows that the selection intensity is largely a function of the proportion of animals selected – the lower the proportion selected, the higher the selection intensity. This is also illustrated in Fig. 4.4. However, with a fixed proportion of animals selected, selection intensities increase slightly as the total number of animals tested rises. For instance, compare the selection intensity when 10 animals are selected from 50 tested (1.372) with that when 20 animals are selected from 100 tested (1.386). As for selection differentials, selection intensities have to be calculated separately for each sex, then averaged.

Table 4.3. Examples of the selection intensity achieved, in standard deviation units, by selecting different numbers of animals from groups of 25, 50, 75 or 100 animals in total. See the table at the end of this book for a fuller set of selection intensities from the same source. (After Becker, 1984.)

Number of animals selected	Total number of animals from which they were selected:			
	25	50	75	100
1	1.965	2.249	2.403	2.508
5	1.345	1.705	1.893	2.018
10	0.936	1.372	1.588	1.730
15	0.624	1.139	1.381	1.536
20	0.336	0.951	1.217	1.386
25	0	0.786	1.079	1.259
35	–	0.488	0.843	1.050
50	–	0	0.539	0.792
75	–	–	0	0.420
100	–	–	–	0

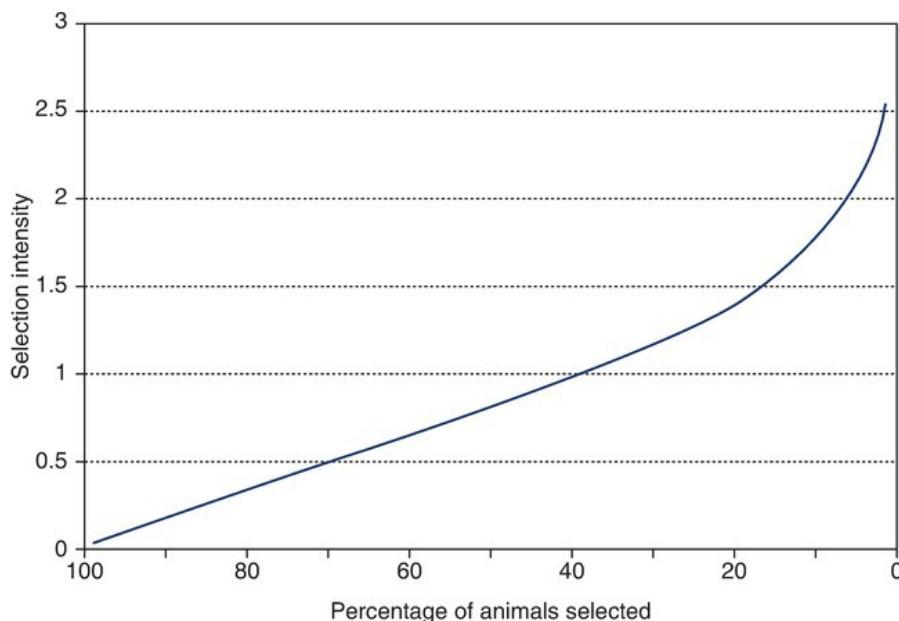


Fig. 4.4. The relationship between the proportion of animals selected and selection intensity – the selection intensity is measured in standard deviation units. The values shown assume that animals are selected from a very large total number of animals tested.

We can calculate the phenotypic variance (VAR_P) or phenotypic standard deviation (sd_P) of the trait under selection in the initial population of animals, as explained in Chapter 2. Alternatively, we could use published estimates of variation from similar animals to those we are interested in, kept in a similar production system (usually variances for the same trait are relatively constant within a given breed and level of production). Putting this together, we can predict selection differentials by multiplying the selection intensity, which tells us how many s.d. units better than average our selected animals are, by the phenotypic s.d., which tells us how much each s.d. unit is worth in kg of live weight, litres of milk, etc. So (after Lush, 1945):

$$S = i \times \text{sd}_P$$

Substituting $i \times \text{sd}_P$ instead of S in the formula used before gives a new formula for predicting response *per annum* (after Lush, 1945):

$$R = \frac{i \times \text{sd}_P \times h^2}{L}$$

It is easiest to see the connection with the earlier formula when the terms are presented as above, but most commonly the terms are re-ordered and the formula is shown as:

$$R = \frac{i \times h^2 \times \text{sd}_P}{L}$$

To illustrate the use of this formula, let us consider a beef cattle breeding programme to improve calf weaning weight (200-day weight) in a herd of 120 Simmental cows. Let us assume that there are 100 calves reared *per annum*, so on average there are 50 calves of each sex available for selection

each year. We estimate that the phenotypic standard deviation of weaning weight in the Simmental breed is 35 kg. If we pick the top five bull calves each year, based on their own weaning weight, then we predict a male selection intensity of 1.705 standard deviations (see Table 4.3). If we select the top 35 heifers each year from 50 recorded, then the female selection intensity will be 0.488 standard deviations (see Table 4.3). This gives an average selection intensity across both sexes of 1.097. If the bulls are mated once only, at 15 months of age, then their progeny will be born when they are 2 years old, and the male generation interval will be 2 years. If the herd is made up of 35 heifers (3 years old), 30 4-year-old cows, 25 5-year-old cows, 20 6-year-old cows and 10 7-year-old cows, and they have equal calving rates, then the weighted average age of dams, and hence the female generation interval, is 5.5 years. The male and female generation intervals together give an average generation interval of 3.25 years. The heritability of weaning weight in beef cattle is usually about 25%. Putting these values into the new formula gives:

$$i = 1.097 \text{ s.d. units}$$

$$sd_p = 35 \text{ kg}$$

$$h^2 = 0.25$$

$$L = 3.25 \text{ years}$$

and so:

$$R = \frac{1.097 \times 35 \times 0.25}{3.25} \text{ kg per annum}$$

$$R = 2.95 \text{ kg per annum (approximately)}$$

The link between selection intensity and generation interval

The formula above shows that the annual response to selection will be highest when the selection intensity is high, the heritability is high, and the generation interval is low. There is little that breeders can do about the heritability of a trait – this is largely a biological characteristic of the trait concerned. But, within biological limits, breeders can increase selection intensities and decrease generation intervals, although these two measures are linked. The strength of that link varies between both species and sexes. For any breeding herd or flock to remain the same size, the breeding animals have to be replaced at least when they die of natural causes or are culled for unavoidable reasons. Because relatively few males are required for breeding, there is a plentiful supply of potential replacement males in all farmed species. This means that it is usually possible to keep male generation intervals short and, at the same time, keep male selection intensities high. But this is not the case for females in all species. The females of most breeds of cattle and sheep rear less than two offspring *per annum*, and so less than one female offspring *per annum* on average. This

automatically sets an upper limit to the number of breeding females which can be replaced each year in these species. This is far less of a constraint in prolific species like pigs, poultry and fish.

It is easiest to see the link between selection intensity and generation interval by considering two extreme examples of a herd or flock replacement policy. At one extreme, all of the available young replacement females could be brought into a herd or flock to replace older females. In this case, the average age of the herd or flock would be low, and female generation intervals would be short. However, there would be no selection at all among female replacements (i.e. the female selection intensity would be 0). At the other extreme, if breeding females were only replaced when they died of natural causes, then the average age of the herd or flock would be high, and the female generation interval would be long. However, in this case few replacement females would be needed, so they could be highly selected, and selection intensities would be high. These two extreme situations, and some of the intermediate ones, are illustrated in Table 4.4.

The optimum flock or herd age structure to maximize response is often intermediate to the two extremes described. This is illustrated in Table 4.5 and Fig. 4.5. The table and graph show the predicted annual rate of response in 20-week weight for the different female replacement rates shown in Table 4.5. In all cases it is assumed that 5 ram lambs are selected from 75 available in each flock each year, and so a male selection intensity of 1.893 standard deviations is achieved. Each ram is used for 1 year only, so the male generation interval is 1 year in all cases. In this example the maximum rate of gain is achieved by replacing 45 females *per annum*, though the difference between alternative schemes is fairly small. The difference between schemes would be more marked if more males were selected, so reducing male selection intensity, or if male generation intervals were longer.

Table 4.4. Female generation intervals and selection intensities achieved in six flocks of 100 ewes when different proportions of available female replacements are brought into each flock. The extreme cases are in flock 1, where females are retained for as long as possible (up to 5 lamb crops), and in flock 6, where all available females are brought into the flock, displacing older ewes. In all cases natural wastage of about 10% is assumed, and ewes are assumed to lamb first at 2 years of age, and to rear an average of 1.5 lambs, of which half are males and half females (for simplicity, it is assumed that ewes of different ages have the same reproductive rate). Selection intensities can be converted to selection differentials by multiplying them by the standard deviation of the trait concerned. For example, if selection is for 20-week weight, the selection intensities would be multiplied by the standard deviation of 20-week weight, assumed to be about 5 kg. So, the female selection differential achieved in flock 1 would be about 5.4 kg ($1.079 \times 5 = 5.395$ kg); that in flock 2 would be about 4.2 kg, and so on.

Ewe age at lambing (years)	Number of ewes in each age group					
	Flock 1	Flock 2	Flock 3	Flock 4	Flock 5	Flock 6
2	25	35	45	55	65	75
3	22	31	40	45	35	25
4	20	27	15	—	—	—
5	18	7	—	—	—	—
6	15	—	—	—	—	—
Female generation interval (weighted average age)	3.76 years	3.06 years	2.70 years	2.45 years	2.35 years	2.25 years
Female selection intensity achieved in 20-week weight (in standard deviation units)	1.079	0.843	0.637	0.443	0.244	0

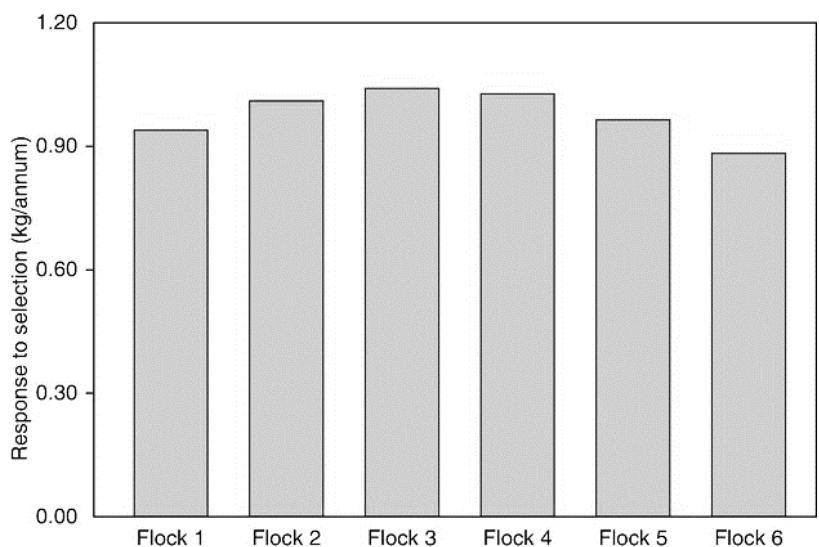


Fig. 4.5. Predicted annual response to selection on 20-week weight in sheep, with different female replacement policies, as shown in Table 4.4.

Table 4.5. Predicted annual response to selection on 20-week weight in sheep, with different female replacement policies. Female generation intervals and selection intensities are those shown in Table 4.4. The heritability of 20-week weight is assumed to be 0.3 and the phenotypic standard deviation of 20-week weight is assumed to be 5 kg. Responses are calculated from the formula shown in the text.

	Flock 1	Flock 2	Flock 3	Flock 4	Flock 5	Flock 6
Female generation interval (years)	3.76	3.06	2.70	2.45	2.35	2.25
Male generation interval (years)	1.00	1.00	1.00	1.00	1.00	1.00
Average generation interval (years)	2.38	2.03	1.85	1.725	1.675	1.625
Female selection intensity	1.079	0.843	0.637	0.443	0.244	0
Male selection intensity	1.893	1.893	1.893	1.893	1.893	1.893
Average selection differential (kg)	1.486	1.368	1.265	1.168	1.069	0.947
Predicted annual response to selection (kg per annum)	0.94	1.01	1.03	1.02	0.96	0.88

In the past, optimizing generation intervals and selection intensities was central to the design of breeding programmes. Deciding on the optimum proportion of animals to select from each age group was important because animals were evaluated usually only once, and the results of these evaluations could only be compared within age groups. The modern methods of genetic evaluation described in Chapter 7 allow comparison of animals of different age groups. Also, advances in the methodology for evaluation and the power of computers means that it is feasible to do evaluations much more often than in the past. When these modern evaluation methods are available, genetic progress will be maximized by selecting animals with the highest predicted breeding value, regardless of age, rather than aiming for a fixed proportion of animals in each age group. The age structure of a flock or herd will still end up close to the optimum calculated as described above. However, there will be some additional gains as a result of keeping animals which are better than expected for their particular age group longer, and culling animals which are poorer than expected for their particular age group sooner.

Using information from relatives

So far, we have considered only a single measurement of performance on the animal itself as the selection criterion. However, this is only one of a number of sources of information which are commonly used in animal breeding. These include records of performance from:

- **The animal's ancestors.** Selection on ancestors' information is called *pedigree selection*. An index combining information on ancestors' performance is usually called a *pedigree index*, and selection on pedigree indexes is common in young animals until they get a record of performance themselves.
- **The animal itself.** Selection on the animal's own performance is the simplest method of selection. Historically, many farm livestock breeding schemes aimed at improving growth have been based on this source of information alone (individual selection or mass selection). When animals' performance is recorded consistently according to a defined protocol (e.g. specific feeding regimes to a target age or weight) this is usually termed a performance test.
- **The animal's full or halfsiblings** (*sibs* for short – these are the animal's brothers and sisters). Full sibs are animals which have both parents in common; half sibs are animals which have only one parent in common. In early population genetics experiments with fruit flies and mice, selection was often practised between families (*family selection*) or within families (*within-family selection*). Family selection involves selection of entire families with the best average performance, from which to breed the next generation of animals. Within-family selection involves selection of the best individual animal in each family, from which to breed the next generation. Family selection can be especially useful in selecting for low heritability traits. However, it can lead to rapid rates of inbreeding, which become difficult to manage with the more constrained sizes of breeding populations of most farmed animal species. However, modifications of family selection were used in early livestock breeding practice and are still useful in establishing breeding programmes for new species. A selection programme based on information from siblings is called *sib selection* or a *sib test*. Sib tests are particularly useful if the trait of interest is only measurable on one sex. For instance, milk records from full sisters are helpful when selecting young dairy bulls. Similarly, early pig breeding programmes intended to improve carcass composition were based on sib selection, following carcass dissection of a full sib. This approach is less common now, because of the widespread use of ultrasonic scanning to assess carcass characteristics in live animals, although it is still used in meat poultry (broiler) breeding. Conversely to the situation in family selection, some form of within-family selection can help to limit rates of inbreeding, and so is often used in conservation of rare breeds, or in maintaining control populations in selection experiments.
- **The animal's progeny.** A breeding programme based on the performance of an animal's offspring is termed a *progeny test*. Again, information on the performance of progeny is particularly useful when the trait of interest can be measured in only one sex. In most countries, selection for milk production in dairy cattle is based on progeny testing. Bulls undergoing tests are used in many milk-recording herds through artificial insemination.

They are then selected for wider use, or culled, on the basis of the milk production records of their daughters, compared with those of *contemporary animals*.

- **Any other relatives of the animal.**
- **Combinations of the classes of relatives listed above.** Modern livestock breeding schemes use information from all available relatives, weighted appropriately, to predict breeding values as accurately as possible. Chapter 7 describes how this is achieved.
- Records of performance from relatives are useful in selection because related animals have genes in common, and so the performance of relatives can provide clues to the genetic merit or breeding value of the candidates for selection. Figure 4.6 shows the proportion of genes in common between the bull marked X, and various classes of relatives. (These proportions are the same for animals of the same relationship in all species.)

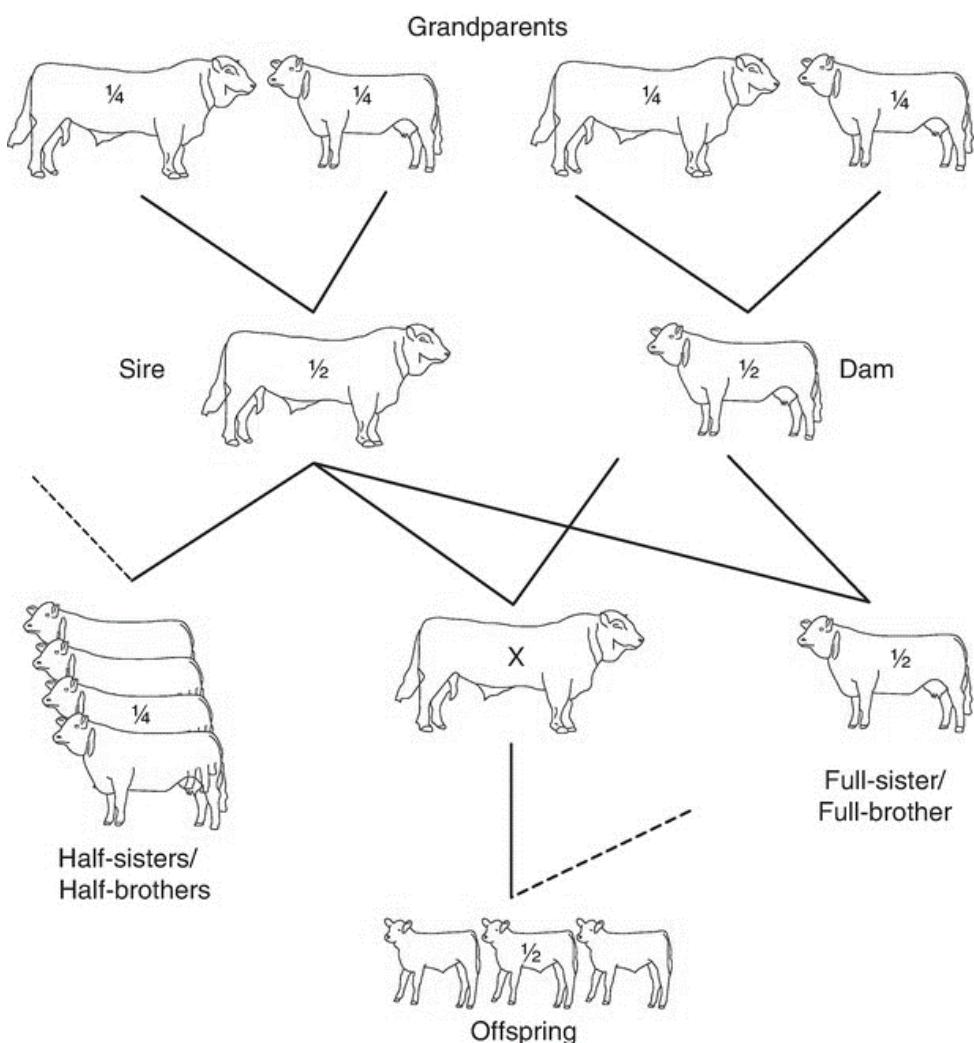


Fig. 4.6. The proportion of genes in common between the bull marked X, and various classes of relatives. Parents and offspring have exactly half their genes in common. The proportions of genes in common between other classes of relatives are expected average proportions. (After Wray and Simm, 1991.)

Offspring always have exactly one-half of their genes in common with each parent (apart from the slight inequalities caused by the different size of the X and Y chromosomes). This is because every animal develops from an embryo that has one member of each pair of its chromosomes from the father, via the sperm, and one member from the mother, via the egg. However, for other classes of relatives the proportion of genes in common are averages. This is because segregation and recombination lead to chance variation in the proportion of genes from ancestors (as explained in Chapter 2). Each sperm and each egg carry copies of half of the genes of the animals that produced them. Which sample of genes make up this half is determined entirely by chance. So, some pairs of full sibs have more than half of their genes in common, by chance, and others have less. (While we have known about this sampling for many years, it has not been easy to measure until recently. However, we can now estimate the proportion of genes in common in any pair of related animals, using the molecular genetic techniques described in Chapter 5.) Each generation, or link in the pedigree diagram, which separates two relatives leads to a halving of the genes they have in common, on average.

In order to predict response to selection using these new sources of information, we need a modified version of the formula used so far:

$$R = \frac{i \times h^2 \times sd_p}{L}$$

For reasons which are explained in the appendix at the end of this chapter, this can be rewritten as (after Falconer and Mackay, 1996):

$$R = \frac{i \times h \times sd_A}{L}$$

In this formula the R, i and L have exactly the same meaning as before. The new terms are h, which is the square root of the heritability, also referred to as the accuracy of selection on a single record of the animal's own performance, and sd_A , which is an abbreviation for the additive genetic standard deviation, or *standard deviation of true breeding values*.

This formula still refers to the special case when selection is on the animal's own performance. There is a more general version in which h is substituted by r, which is the accuracy of selection on any combination of records from the animal and its relatives:

$$R = \frac{i \times r \times sd_A}{L}$$

As mentioned in Chapter 2, the symbol r is usually used to denote a correlation. In this case the accuracy is the correlation between animals' true breeding values for the trait(s) under selection, and the measurement(s) on which selection is based.

Accuracy of selection

The accuracy of selection itself depends mainly on three things: (i) the heritability of the trait concerned – the higher the heritability, the higher the accuracy; (ii) the source of information on which selection is based, e.g. the class of relative – the closer the relatives on whose records selection is based, the higher the accuracy of selection; and (iii) the amount of information available from relatives – the more relatives of a given class recorded (and the more often they are recorded, if the trait can be measured repeatedly), the higher the accuracy, although there are diminishing returns as the numbers of relatives increase. Table 4.6 shows the relationship between accuracy and the square root of the heritability, for different classes of relatives, assuming that selection is based on a single record of performance from that type of relative. To predict response to selection on such a record, we simply use the appropriate value of r in the formula above. So, for example, if selection is on a single record on the animal itself, the accuracy of selection is equal to the square root of the heritability of the trait concerned ($r = h$). To continue an example used earlier, the heritability of weaning weight in beef cattle is about 0.25, so its square root, h , is 0.5. Hence the accuracy of selection for weaning weight, when selection is based on a single record of the animal's own performance, is 0.5, or 50%. If selection is based on a single measurement from one parent, or one offspring, then the accuracy is only half that when selection is based on a record on the animal itself. In the beef cattle example, selection on a single parent or offspring record would give an accuracy of 0.25 or 25%.

Table 4.6 shows that the accuracy of selection on a single record of performance for different types of relative is directly related to the proportion of genes they have in common. This is a very important principle in objective animal breeding. In subjective selection, it is easy to overplay the importance of a particular favourite ancestor, or other relative, in making selection decisions (e.g. if it has won prizes in major shows or was sold for a very high price). In objective selection, the aim is to give each performance record from a relative its proper emphasis, and this emphasis depends on the expected proportion of genes in common between the relative and the candidate for selection.

Calculating the accuracy of selection with records from greater numbers of relatives is slightly more complicated, although the results are easy to understand. Figure 4.7 shows the accuracy of selection on records from varying numbers of half or full sibs and progeny.

Table 4.6. The accuracy of selection on a single record of performance on the candidate animal itself, or on a single record from a relative.

Accuracies are presented as proportions of h , the square root of the heritability of the trait under selection. So, for example, if $h^2 = 0.25$, $h = 0.5$, the accuracy of selection on the animal's own performance is 0.5, the accuracy of selection on one parent's record, or one full sib's record, or one progeny's record is 0.25 ($\frac{1}{2} \times 0.5 = 0.25$), the accuracy of selection on one grandparent's record, or one half sib's record, is 0.125 ($\frac{1}{4} \times 0.5 = 0.125$), and so on. (After Nicholas, 1987.)

Single measurement of performance on:	Accuracy (r)
Animal itself	h
One parent	$\frac{1}{2} h$
One grandparent	$\frac{1}{4} h$
One great-grandparent	$\frac{1}{8} h$
An identical twin	h
One full sib	$\frac{1}{2} h$
One half sib	$\frac{1}{4} h$
One progeny	$\frac{1}{2} h$

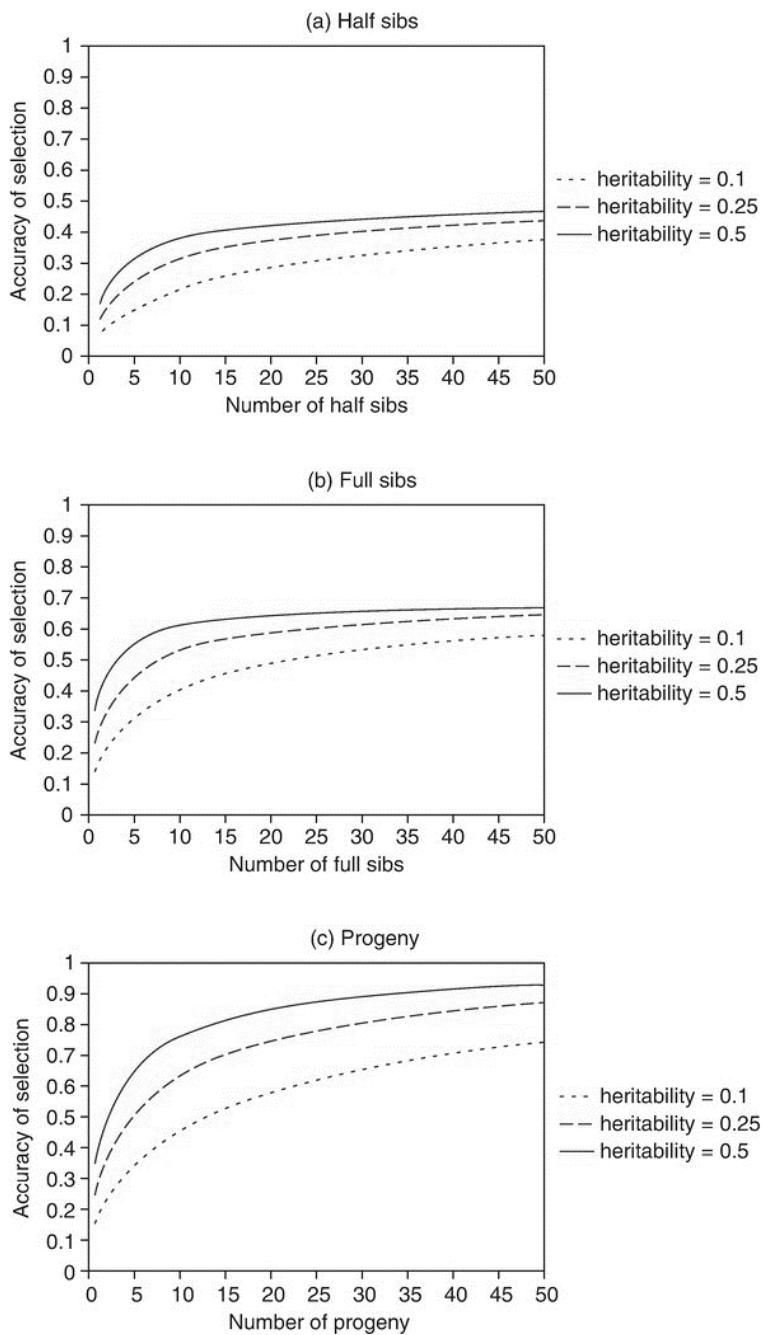


Fig. 4.7. The accuracy of selection on performance records from relatives with varying numbers of records, and traits of different heritability. In graph (a) the records are from half sibs, in graph (b) they are from full sibs, and in graph (c) they are from progeny. (After Nicholas, 1987.)

These graphs illustrate that:

- The closer the relative, the more valuable the record. For example, a single record from an offspring is twice as valuable as a single record from a half sib, other things being equal (i.e. the accuracy of selection is twice as high).

- As the number of progeny increases, the accuracy of selection tends towards 1. As the number of full sibs increases, the accuracy tends towards 0.7. With half sibs, the accuracy tends towards 0.5 as numbers increase.
- A single record from a full sib is of equal value to a single progeny record, but thereafter records from progeny are more valuable in terms of increasing the accuracy of selection.
- Initially, the more relatives of a given class, the higher the accuracy, although there are diminishing returns.
- Initially, the higher the heritability, the higher the accuracy.
- Initially, the lower the heritability, the greater the proportional contribution which records from relatives make (e.g. compared to the accuracy of selection on a single record on the animal itself, selection on records from 50 progeny increases the accuracy of selection by a factor of about 5 when the heritability is 0.1, but only by a factor of about 1.6 when the heritability is 0.3).

Additive genetic standard deviation

The value of the additive genetic standard deviation or standard deviation of breeding values can be derived from the heritability of the trait concerned, and its phenotypic standard deviation. We have already seen that:

$$\text{Heritability } (h^2) = \frac{\text{VAR}_A}{\text{VAR}_P}$$

Standard deviations are the square root of variances, and so we can take the square root of all terms in this equation to get:

$$h = \frac{\text{sd}_A}{\text{sd}_P}$$

We can re-arrange the terms in this equation (by multiplying both sides by sd_P , then cancelling out the two sd_P s which appear on the right-hand side of the equation) to give:

$$\text{sd}_A = h \times \text{sd}_P$$

So, the additive genetic standard deviation is simply the phenotypic standard deviation multiplied by the square root of the heritability.

A practical example

Although using records of performance from relatives can increase the accuracy of selection, it is important to recognize that this benefit can be outweighed by increases in the generation interval. For example, if we want to select for increased 400-day weight in a 120-cow herd of beef cattle, we might consider two main options, performance testing and progeny testing. For simplicity, let us assume that selection is based on males only, and that all available replacement females enter the herd, giving a female generation interval of 3.85 years and a female selection intensity of 0. Also, let us assume that progeny testing takes place in several separate crossbred herds. In each case, if we select 5 males from a total of 50 which are either performance recorded in the pure herd, or progeny tested in the crossbred herds, then the selection intensity among males is 1.705. The average selection intensity is then 0.8525 ($[1.705 + 0]/2$).

The heritability of 400-day weight is about 0.4, so if we select bulls on their own performance alone, the accuracy of selection is about 0.632 (the square root of the heritability). Subject to them reaching sexual maturity, we should be able to mate bulls at about 15 months of age, giving a male generation interval of 2 years. This means that in the herd using performance testing the average generation interval is 2.925 years ($[3.85 + 2]/2$). If we select bulls on the records from 20 progeny, the accuracy is higher – about 0.830 – but the crossbred progeny do not reach 400 days of age until the bulls are over 3 years old. So, the minimum male generation interval possible in the purebred herd is approaching 4 years (it is safest to assume 4 years exactly if we want to maintain a seasonal pattern of calving). This means that in the herd using progeny testing the average generation interval is 3.925 years ($[3.85 + 4]/2$).

To predict annual response to selection in the two schemes, we also need to know the additive genetic standard deviation of 400-day weight. This is about 28.4 kg for the larger beef breeds. (This is calculated from the phenotypic standard deviation of 400-day weight (about 45 kg) multiplied by the square root of the heritability of 400-day weight (square root of 0.4 = 0.632); so $45 \times 0.632 = 28.4$ kg. Table 4.7 summarizes the characteristics of the two breeding schemes.

	Performance testing	Progeny testing
Average selection intensity (i)	0.8525	0.8525
Approx. accuracy of selection (r)	0.632	0.830
Additive genetic standard deviation (sd_A in kg)	28.4	28.4
Average generation interval (L in years)	2.925	3.925
Predicted response (kg per annum, approx.)	5.23	5.12

Table 4.7. Characteristics of the two beef cattle breeding schemes described in the text – one based on performance testing, the other on progeny testing. In both cases selection is to increase 400-day weight.

The predicted annual response to selection can then be calculated from these characteristics, using the formula introduced in the last section. This is illustrated below for the scheme based on performance testing:

$$i = 0.8525$$

$$r = 0.632$$

$$sd_A = 28.4 \text{ kg}$$

$$L = 2.925$$

and so:

$$R = \frac{0.8525 \times 0.632 \times 28.4}{2.925}$$

$$R = 5.23 \text{ kg per annum (approx.)}.$$

The equivalent response for the scheme based on progeny testing is 5.12 kg *per annum*. In this example, the gains in accuracy from progeny testing are outweighed by the longer generation interval among males which is necessary to obtain records of performance on progeny. However, with a lower heritability, or more progeny records, the reverse could be true. So, it pays to check the expected responses from different types of selection, using the relevant values for the heritability and number of relatives to derive the accuracy, before embarking on selection. Usually records from ancestors and at least some sibs (or other *collateral relatives* – those from the same generation as the animal itself) can be used to increase accuracy of selection without lengthening the generation interval. Using records from descendants usually increases the generation interval, and so there is a trade-off between increasing accuracy and increasing generation interval. The methods of combining information from different types of relative are discussed in later sections. As we will see later, genomic information is revolutionizing animal breeding, by increasing the accuracy of selection with the same, or shorter, generation intervals.

Comparisons of different breeding schemes also need to consider the resources needed, as well as predicted responses. Even when it appears justified by higher predicted responses, it would be very expensive to maintain crossbred herds solely for progeny testing. In some cases, equivalent responses could be achieved at lower cost from performance testing, by expanding the purebred herd to increase selection intensity. Also, in most cases it would be quite wasteful to progeny test all available bulls. If the number tested is reduced arbitrarily, then this will lead to lower selection intensity than that possible from performance testing. In practice, it makes sense to use all available information in selection – in the example above it would have been more efficient to progeny test only those bulls which themselves had high 400-day weight (a process called two-stage selection). Also, records from half sibs are generated as a by-product of most performance testing schemes, so

these can often increase accuracy at no extra cost in money or time. (See Weller (1994) for methods of comparing the cost-effectiveness of different breeding schemes.)

Using repeated records of performance

So far, we have considered only single records of performance on candidates for selection, or their relatives. However, many traits of interest can be measured more than once over an animal's lifetime. For example, growing meat animals can be weighed on several occasions, most dairy cows have about four lactations during their lifetime, breeding pigs usually have multiple litters of piglets during their lifetime, most beef cows would wean six or seven calves during their lifetime, and sheep would have four or five litters during their lifetime and be shorn annually for four or five years.

Often there are very strong genetic associations between performance at different stages in an animal's life. In these cases, it is safe to assume that performance at these different stages is being influenced by the same genes. Hence, repeated records of performance can give additional clues to the breeding value of animals and enhance the accuracy of selection. If repeated measurements are influenced by exactly the same genes, then any differences in performance from contemporaries, for example in yield in different lactations, is due to differences in the environment or management experienced by the animal. We have already seen that the performance of an animal can be split into genetic and environmental components (after Falconer and Mackay, 1996):

$$\text{Phenotype (P)} = \text{Genotype (G)} + \text{Environment (E)}$$

Splitting the genetic part down further into additive genetic merit (breeding value), plus non-additive genetic effects, as explained in Chapter 2, gives:

$$\begin{aligned}\text{Phenotype (P)} &= \text{Additive genetic (A)} + \\ &\quad \text{Non-additive genetic (NA)} + \\ &\quad \text{Environment (E)}\end{aligned}$$

However, the environmental component of performance can also be split further into permanent (E_p) and temporary (E_t) environmental effects. This gives:

$$P = A + NA + E_p + E_t$$

Like the genes, *permanent environmental effects* stay with the animal for life but, unlike the genes, they do not get passed on to offspring. *Temporary environmental effects* have a much shorter impact, for example affecting only one lactation or one weight record. If a dairy heifer suffered a severe mastitis infection in one-quarter of her udder, and as a result of damage to the secretory tissue in that quarter, never produced milk from it again, that would be a permanent environmental effect. Yield in the other quarters may rise slightly to compensate, but the animal would produce less milk in total in each successive lactation, than if it had not been affected by mastitis. In a single lactation, the yield of the same animal may be temporarily depressed because it happened to be housed in a

group of cows which was fed very poor-quality silage. However, when the animal was fed on better silage in subsequent lactations, this effect would disappear. So, this would be classified as a temporary environmental effect.

The value of repeated records in selection depends on a measure called the *repeatability*. This is the correlation between repeated records from the same animal. It is defined as the proportion of the total phenotypic variation which is explained by the combined effects of the genes and the permanent environmental variation (after Falconer and Mackay, 1996):

$$\text{Repeatability} = \frac{\text{VAR}_G + \text{VAR}_{Ep}}{\text{VAR}_P}$$

or, in full:

$$\text{Repeatability} = \frac{\text{VAR}_A + \text{VAR}_{NA} + \text{VAR}_{Ep}}{\text{VAR}_P}$$

This looks similar to the equation used earlier to calculate the heritability, except that there are a few extra terms on the top of the equation. The heritability is concerned only with the 'predictable' additive genetic part of the equation, whereas the repeatability also includes the variation due to non-additive genetic and permanent environmental effects, which are specific to individual animals, and hence difficult to predict. As a result, the repeatability sets an upper limit to the heritability. Like heritabilities, repeatabilities range from 0 to 1, or from 0 to 100%. Some examples of the repeatability of different traits are shown in Table 4.8.

It may seem counter-intuitive at first sight, but the higher the repeatability of a trait, the lower the value of repeated records in selection for that trait. In other words, if the repeatability is high, the first record of performance gives a good indication of the animal's genetic merit, and adding subsequent records only improves the accuracy slightly. If the repeatability is low, the first record of performance gives a poor indication of subsequent performance. Here, repeated records help to build up a more accurate estimate of the animal's genetic merit, but the law of diminishing returns applies. This is illustrated in Fig. 4.8, which shows the accuracy of selection for traits with different repeatability, depending on the number of repeated records available. As with the use of information from relatives,

Table 4.8. Estimates of the repeatabilities of some traits of economic importance in livestock.

	Flock 1	Flock 2	Flock 3	Flock 4	Flock 5	Flock 6
Female generation interval (years)	3.76	3.06	2.70	2.45	2.35	2.25
Male generation interval (years)	1.00	1.00	1.00	1.00	1.00	1.00
Average generation interval (years)	2.38	2.03	1.85	1.725	1.675	1.625
Female selection intensity	1.079	0.843	0.637	0.443	0.244	0
Male selection intensity	1.893	1.893	1.893	1.893	1.893	1.893
Average selection differential (kg)	1.486	1.368	1.265	1.168	1.069	0.947
Predicted annual response to selection (kg per annum)	0.94	1.01	1.03	1.02	0.96	0.88

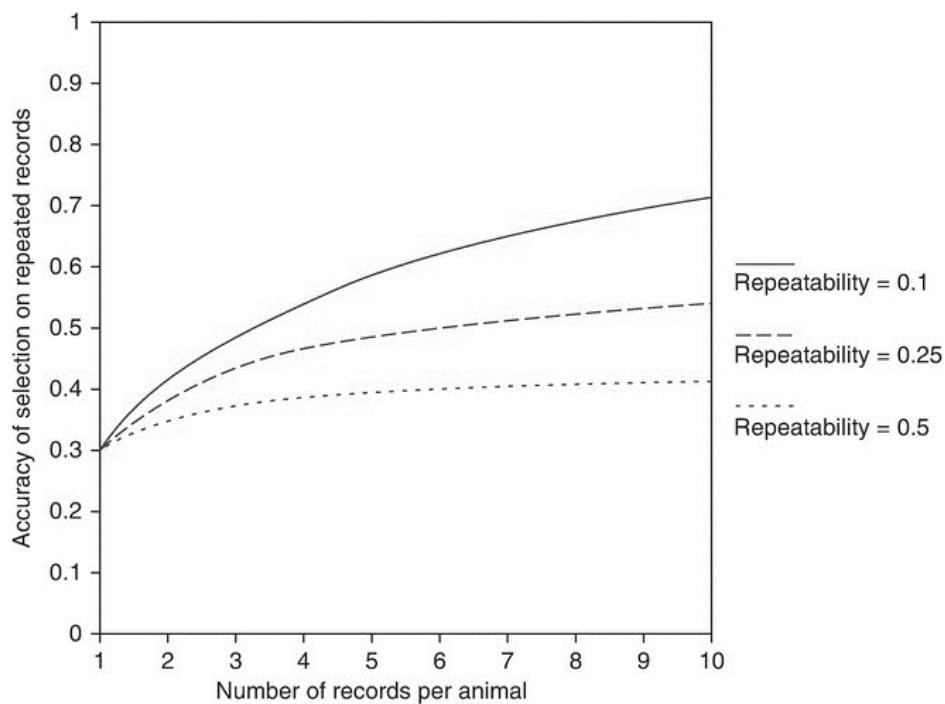


Fig. 4.8. The effect of different numbers of repeated records on the accuracy of selection, for traits with different repeatabilities. In this example the heritability of the trait is 0.1.
(After Nicholas, 1987.)

it is important to take into account any time lag in obtaining repeated records. In principle, the benefits of higher accuracy are often outweighed by the penalty from increases in generation interval. However, in practice, selection is usually based on a single record initially, with subsequent records being used to refine later estimates of genetic merit.

Repeatabilities measure the similarity between repeated records of the same trait, assuming that performance measured in different years (or at other intervals) is controlled by exactly the same genes. However, there are often associations between different traits which are not likely to be influenced by exactly the same genes. For instance, animals that are heavier than their contemporaries at a given age tend to be fatter too; and higher yielding cows generally have lower milk protein percentage. Even when the same trait is measured at different times, different genes may influence the trait at different times. For example, animals which are heavier than average at an early age are often heavier at later ages, but not all the genes influencing early weight also influence mature weight. Similarly, heifers which have a high milk yield compared to contemporaries, often have comparatively high yields in later lactations, and *vice versa*, but not all the genes which influence early lactations also influence later lactations. The degree of association between different characteristics is measured by the correlation coefficient, as explained in Chapter 2.

Predicting correlated responses to selection

Before embarking on a selection programme, it is useful to predict what the consequences will be, not only in the trait under selection, but also in other associated traits. Providing that the heritabilities of the trait under selection and other traits of interest are known, their phenotypic variances (or standard deviations) are known, and the genetic correlation between them is known, the *correlated response* to selection can be predicted. It is easiest to think of this as a three-step procedure:

1. Calculate the predicted annual response in the trait under selection (trait X). If we want to predict correlated responses it is simplest if we first predict the direct response in the trait under selection *in additive genetic standard deviation units*, rather than units of measurement. If selection is based on a single record of the animal's own performance, this becomes:

$$R_X = \frac{i \times h_X}{L} \text{ (in additive genetic s.d. units of X per annum)}$$

(Note that this formula is identical to the one presented earlier, except that the term s_{dA} is omitted, in order to express the response in additive genetic s.d. units, rather than units of measurement.)

2. Multiply the predicted annual response from (1) by the genetic correlation between the traits under selection and the second trait of interest (trait Y), to get the predicted correlated response in the second trait, *expressed in additive genetic standard deviation units*:
 $CR_Y = R_X \times r_{AXY}$ (in additive genetic s.d. units of Y per annum)

3. Multiply the predicted correlated response in standard deviation units from (2) by the additive genetic standard deviation of the second trait (s_{dAY}) to express the predicted correlated response *in the units of measurement per annum*:

$$CR_Y = R_X \times r_{AXY} \times s_{dAY} \text{ (in units of measurement per annum)}$$

Writing this out in full gives (after Falconer and Mackay, 1996):

$$CR_Y = \frac{i \times h_X \times r_{AXY} \times s_{dAY}}{L}$$

To extend an example used before, if we select for 20-week weight in sheep, we expect an increase in fat depth as a result of a positive genetic correlation between these two traits. We can predict the direct response in weight (in units of measurement) from one of the formulae presented earlier in this chapter, and the correlated response in fat depth using the formula above. If we assume that:

$$i = 1.4$$

$$h^2_x = 0.3$$

$$h_x = 0.55$$

$$h^2_y = 0.3$$

$$h_y = 0.55$$

$$r_A = 0.4$$

$$sd_{AX} = 2.75$$

$$sd_{AY} = 0.715$$

$$L = 2.0$$

then the predicted response in 20-week weight is:

$$R_x = \frac{i \times h_x \times sd_{AX}}{L} \quad (\text{in kg per annum})$$

$$R_x = \frac{1.4 \times 0.55 \times 2.75}{2.0}$$
$$R_x = 1.06 \text{ kg per annum (approx.)}$$

and the correlated response in fat depth is:

$$CR_y = \frac{1.4 \times 0.55 \times 0.4 \times 0.715}{2.0}$$

$$CR_y = 0.11 \text{ mm per annum.}$$

Limitations of these predictions

Reduction in variation following selection

There are several limitations to the methods of predicting response described above which need to be mentioned. The first concerns the amount of variation in the trait under selection. Selection itself causes a reduction in the amount of genetic variation in the first few generations. In other words, choosing parents with performance at one end of the distribution results in a narrower distribution in the performance and breeding values of their offspring. The reduction in variation depends on the heritability of the trait concerned and the intensity of selection. The higher the heritability, and the higher the selection intensity, the greater the reduction in variation from one generation to the next. As a result of this, predictions of long-term responses to selection based on the initial variation in the population may be too high. For example, with intense selection for a trait with a high heritability, there may be a loss of about 20% in response to the second generation of selection, compared to that obtained in the first generation (Falconer and Mackay, 1996). This reduction in genetic variation happens in the first few generations of selection, and then the amount of variation stabilizes, or approaches equilibrium. Typically, the loss in response to selection compared to that in the initial generation, stabilizes at about 25%. There are two ways to deal with this problem. The first is to re-estimate the amount of variation and the heritability after the first few generations of selection, and then re-calculate predicted responses from these equilibrium values. The second, which is less costly, is to use more complex formulae to predict long-term response to selection, which account for the fact that heritability and genetic variation were estimated from populations in which selection had not been practised (Dekkers, 1992; Villanueva *et al.*, 1993).

Selection limits

After many generations of selection, the variation in the trait of interest may become exhausted – for example, if each animal in the population has copies of only those alleles that have a favourable effect on performance. As a result, no further response to selection can be achieved. While these *selection limits* or *plateaux* have been reached in some selection experiments with fruit flies and other laboratory species, they are rarely of practical concern in farm animals (see Simm *et al.* (2004) for a review). This is because they occur after many generations of selection, which implies many decades of selection for the same trait in farm livestock. Over this time span it is unusual for market requirements to remain the same, and so it is unlikely that selection is for exactly the same animal characteristics. Also, livestock populations are rarely closed over this length of time – the introduction of new animals introduces new variation and hence delays the approach to the selection limit. Even in a closed population, with selection for the same trait, new mutations prolong the period over which responses to selection are achieved. So, in theory, predictions of response do not hold true indefinitely. But, in practice, they are usually satisfactory over the lifetime of a particular breeding programme.

Variation in response

The final limitation which needs to be mentioned concerns the variation in response that may be achieved in different selection programmes. The formulae presented earlier predict the average response to selection. However, if several identical breeding programmes were started at the same time, they would not achieve identical responses. Some schemes would achieve lower responses than predicted, others would achieve higher responses, but on average the response should match that predicted. It is usually too expensive to demonstrate this variation in response by running identical experimental breeding programmes in farm livestock, but there are many good examples of this in laboratory animal selection experiments. However, for fairly simple breeding programmes, it is possible to predict the variation in response.

The variation in the response achieved occurs because there is an element of chance as to which genes are present in the initial group of animals. There is also an element of chance in which genes are lost and which ones are retained each time parents are selected. (These elements of chance are fundamental to the study of genetic improvement of livestock and help to make it such a fascinating subject for scientists and practical breeders alike.) The main factors affecting the variation in response from that predicted are: (i) the size of the population – the smaller the population, the greater the risk of a difference between predicted and achieved responses; and (ii) the number of parents selected each generation – the smaller the number of parents selected, particularly males as there are usually fewer of these needed in the first place, the greater the risk of a difference between predicted and achieved responses.

Controlling Inbreeding

As explained earlier, inbreeding is the mating of two related animals. Sometimes breeders pursue a deliberate policy of mating related animals, with the aim of increasing the frequency of favourable genes. However, even when steps are taken to avoid it, inbreeding is inevitable in closed populations. Sooner or later, matings will occur between related animals. In smaller populations this will tend to occur sooner, and in larger populations it will tend to occur later. The chances of related animals mating are increased when selection is practised in a closed population. This is because related animals have genes in common and hence their performance and breeding values are more alike than those of unrelated animals. As a result, selection for a particular characteristic increases the frequency of matings between related animals and hence increases inbreeding compared to that in a random mating population.

There are two main reasons for wishing to limit inbreeding. The first is that inbreeding reduces the amount of genetic variation in a population, and so can reduce response to selection (and increase variation in response to selection). The second is that inbreeding can lead to a decline in performance in traits associated with fitness, such as reproductive rate and disease resistance. This decline in performance is known as *inbreeding depression*. Table 4.9 gives some examples of inbreeding depression in a range of characteristics in farm livestock. Inbreeding depression is thought to be the result of an increase in the frequency of recessive genes which adversely affect those characters associated with survival and overall 'fitness'. These are generally the same sort of traits

Table 4.9. Examples of inbreeding depression in traits of economic importance.

Animals	Breed	Trait	Inbreeding depression expressed as change in trait per 1% increase in inbreeding	Source
Dairy cattle	Holstein	Milk yield	-25 kg (-0.37%) ^a	Miglior <i>et al.</i> (1995b)
		Fat yield	-0.9 kg (-0.35%)	Miglior <i>et al.</i> (1995b)
		Protein yield	-0.8 kg (-0.36%)	Miglior <i>et al.</i> (1995b)
		Fat %	+0.05% (+1.32%)	Miglior <i>et al.</i> (1995b)
		Protein %	+0.05% (+1.55%)	Miglior <i>et al.</i> (1995b)
		Somatic cell score	+0.012 (+0.51%)	Miglior <i>et al.</i> (1995a)
Beef cattle	Hereford	Pregnancy rate (2 y.o. cows)	-0.23% (-0.33%) ^b	MacNeil <i>et al.</i> (1989)
		Prenatal survival (2 y.o.)	-1.67% (-1.73%)	MacNeil <i>et al.</i> (1989)
		Weaning rate (2 y.o.)	-1.24% (2.26%)	MacNeil <i>et al.</i> (1989)
		Weaning weight (as trait of cow; all ages)	-0.47 kg (0.24%)	MacNeil <i>et al.</i> (1989)
Sheep	S. Blackface, Cheviot, W. Mountain	Gross lifetime income per ewe	-£1.27 (1.16%) ^b	Wiener <i>et al.</i> (1994a)
		Greasy fleece weight	-0.01 kg (0.53%)	Wiener <i>et al.</i> (1994b)
	Many breeds	Greasy fleece weight	-0.017 kg	Lamberson and Thomas (1984)
		Weaning weight	-0.111 kg	Lamberson and Thomas (1984)
		Ewe fertility (ewes lambing/ewe joined)	-0.014	Lamberson and Thomas (1984)
		Lamb survival (lambs weaned/lambs born)	-0.028 ^c / -0.012 ^d	Lamberson and Thomas (1984)
		Total number piglets born	-0.012 to -0.023	Saura <i>et al.</i> (2015)
Pigs	Iberian (Guadyerbas strain)	Total number piglets born alive	-0.012 to -0.023	Saura <i>et al.</i> (2015)
		Age at first egg	+0.83 to +1.08%	Sewalem <i>et al.</i> (1999)
Chickens	White Leghorn (lines selected for egg number, weight or mass)	Egg number at 63 weeks	-1.67 to -2.75%	Sewalem <i>et al.</i> (1999)
		Egg weight at 63 weeks	0 to -0.94%	Sewalem <i>et al.</i> (1999)

^a All figures in brackets express inbreeding depression as a percentage of the mean of the trait concerned when the mean was presented;

^b figures in brackets express inbreeding depression as a percentage of the mean performance of crosses between inbred and control lines;

^c per 1% increase in inbreeding of lamb;

^d per 1% increase in inbreeding of dam.

which show heterosis as a result of crossbreeding. So, it is helpful to think of inbreeding and inbreeding depression as being the opposite of crossbreeding and heterosis, respectively.

The decline in genetic variation following inbreeding, and the amount of inbreeding depression, both depend on the amount of inbreeding. The closer the relationship between two animals which are mated, the greater the amount of inbreeding in the resulting offspring. Hence, it is important to be able to measure the amount of inbreeding of existing animals, as well as the animals which could be produced from a particular future mating. It is also important to be able to predict the rate at which inbreeding will accumulate over time in a herd, flock or breed involved in a particular breeding programme.

The amount of inbreeding is measured by the *inbreeding coefficient* (abbreviated F). The inbreeding coefficient is formally defined as the probability that two alleles at any locus are *identical by descent*. This is not as complicated as it sounds. Consider two calves, both of which inherit two genes for black coat colour (BB). The first calf is the result of a mating between an unrelated bull and cow. It has two alleles which are *identical by state* (i.e. they are both B alleles, and they have the same effect, but they originated in unrelated ancestors). The second calf is the result of a mating between two heterozygous black half sibs (Bb). The pedigree of the second calf is shown in Fig. 4.9. In this example the bull and cow that are mated are paternal half sibs, i.e. they have the same sire (the calf's grandsire). If this is a heterozygous black bull (Bb), and the calf's granddams are both homozygous red cows (bb), then the calf must have obtained both copies of his B allele from the common grandsire. In this case the two B alleles are identical by descent, i.e. they are copies of a single allele in a common ancestor. The inheritance of segments of chromosome that are identical by descent (IBD) is illustrated in Fig. 4.10.

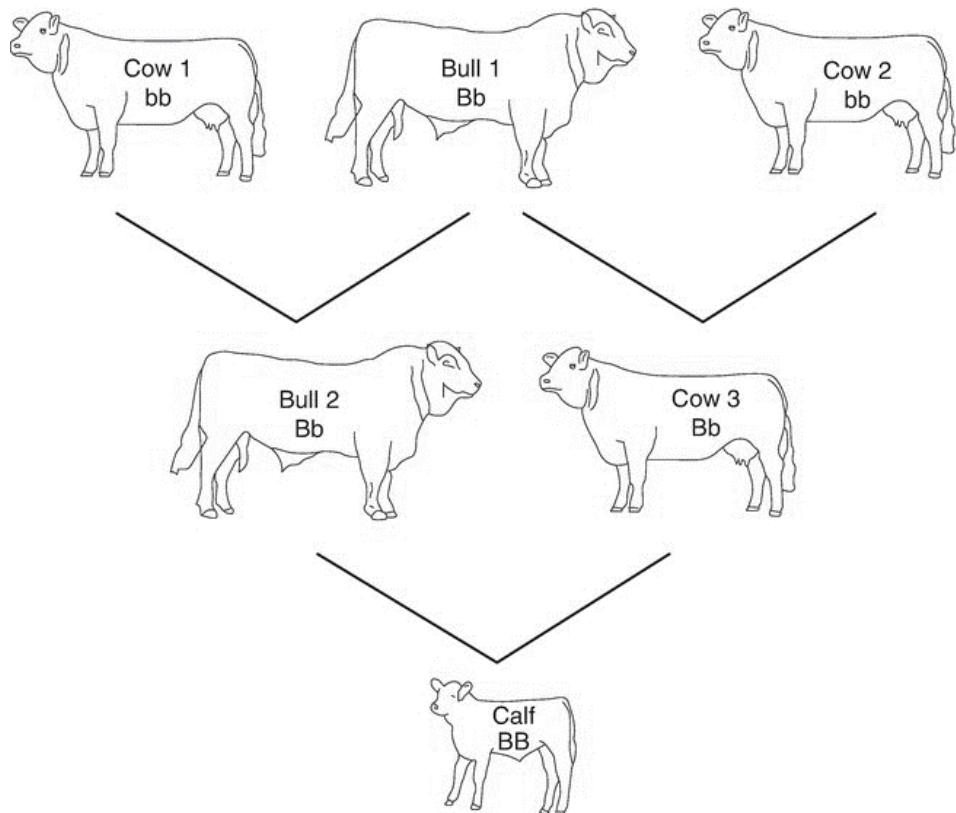


Fig. 4.9. An illustration of how an animal inherits two alleles which are *identical by descent*, i.e. they are copies of the same allele from a single ancestor. In this example Bull 1 with genotype Bb is mated to Cows 1 and 2, both with genotype bb. A son of Cow 1 (Bull 2) with genotype Bb is then mated to his half-sister (Cow 3) also of genotype Bb. The resulting calf has the genotype BB. This calf must have inherited both copies of the B allele from Bull 1, as this bull is the only animal in the grandparent generation to carry a copy of this allele.

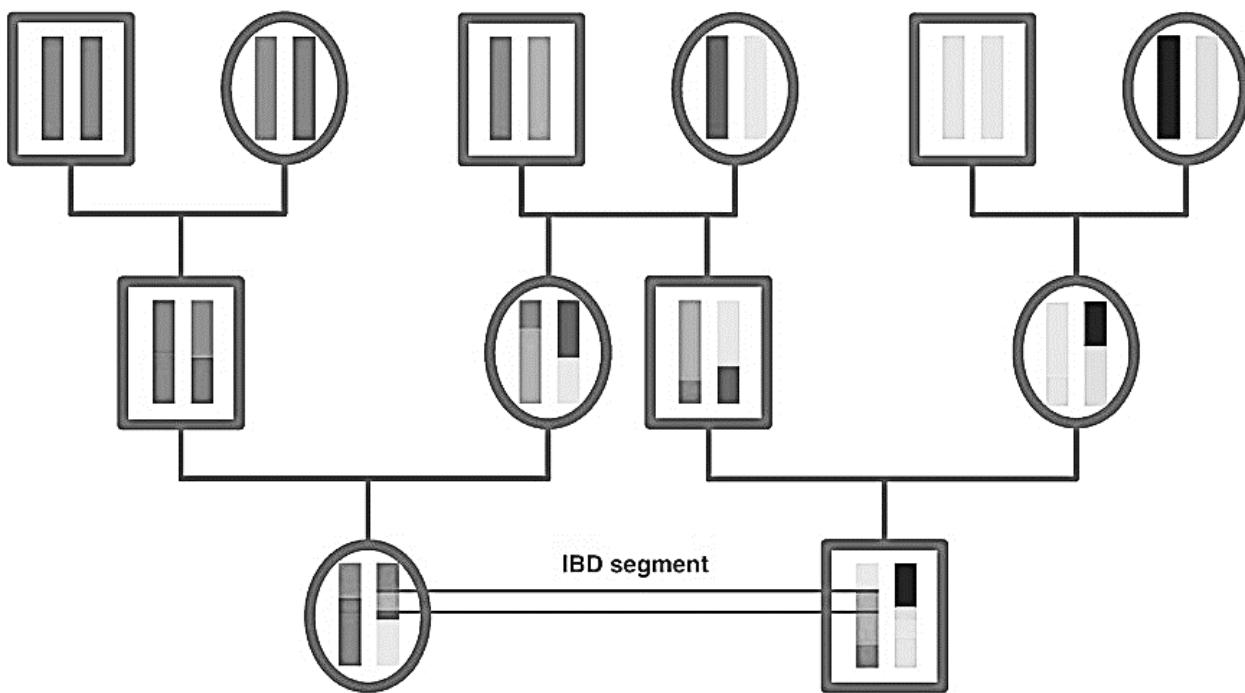


Fig. 4.10. The inheritance of segments of chromosome that are identical by descent (IBD). (Available at: https://commons.wikimedia.org/wiki/File:Pedigree,_recombination_and_resulting_IBD_segments,_schematic_representation.png, by Gklambauer, CC-BY-SA 3.0, accessed 8 December 2019.)

Figure 4.11 shows an abbreviated version of a pedigree of an animal which is the result of a mating between two half sibs (such as the calf in the previous example). The inbreeding coefficient of the animal at the bottom of the pedigree (animal D) can be calculated from pedigrees like this by counting the number of individuals in each path through the pedigree which leads from the animal whose inbreeding coefficient is being calculated, through a common ancestor and back to the original animal. This is easiest to follow by referring to Fig. 4.11. If we start with animal D, we can trace a path through B to the common ancestor A and back through C to D again. In other words, there are three animals in the common path excluding D itself – these are B, A and C. If there is more than one common ancestor, then this process is repeated for each one. The inbreeding coefficient can be calculated from the formula (after Lush, 1945):

$$F_D = \sum (\frac{1}{2})^n (1 + F_A)$$

where:

- the symbol Σ indicates that the results are summed for each path through a common ancestor;
- n is the number of individuals in each path (the $\frac{1}{2}$ appears in the formula to account for the fact that only 50% of sperm or eggs get a copy of the allele being traced from the common ancestor; the other 50% will get a copy of an allele from other unrelated ancestors). The formula shows $\frac{1}{2}$ ‘raised to the power of n’ – so if n = 2, the $\frac{1}{2}$ is squared ($\frac{1}{2} \times \frac{1}{2}$), if n = 3, the $\frac{1}{2}$ is cubed ($\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2}$), and so on;

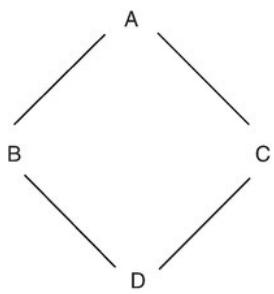


Fig. 4.11. An abbreviated and more general version of the pedigree shown in Fig. 4.9, where two half sibs are mated together. Animal A is the common ancestor, i.e. the father of both animals B and C if they are paternal half sibs, or their mother if they are maternal half sibs. Animals B and C are of opposite sex and they are mated to produce animal D.

- F_D is the inbreeding coefficient of animal D; and
- F_A is the inbreeding coefficient of animal A.

If the inbreeding coefficient of animal A is unknown, then it is assumed to be 0, and the formula simplifies to:

$$F_D = \sum(\frac{1}{2})^n$$

So, in the example above where there is only a single common ancestor, and three individuals in the path, the inbreeding coefficient of animal D is calculated from:

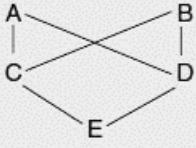
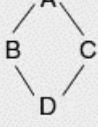
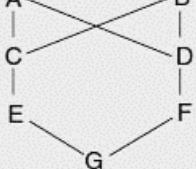
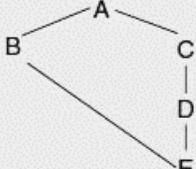
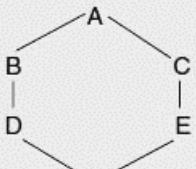
$$F_D = (\frac{1}{2})^3$$

$$F_D = 0.125 \text{ or } 12.5\%.$$

In other words, there is a 12.5% chance that the two alleles at any locus are IBD in an animal which results from a mating between half sibs. This example also illustrates the fact that, generally, the inbreeding coefficient of an animal is one-half of the additive relationship between its parents. In this case the parents were half sibs, which are expected to have 25% of their genes in common, so they are said to have an additive relationship of 25%. So, when half sibs are mated to each other we expect their offspring to have inbreeding coefficients of 12.5%. Table 4.10 shows pedigrees and inbreeding coefficients for some other types of mating between close relatives.

If pedigrees could be traced back far enough, then all animals in a species would be related to each other, though the relationship would be very distant in most cases. It is not practical, or necessary, to go this far back in calculating inbreeding coefficients in livestock breeding, but it is important to define the reference population from which coefficients are calculated. For example, we might define the first generation of animals recorded in a herd or flock book, or animals born in some other significant year, as the initial reference population. Then, in calculating inbreeding coefficients, we assume that none of the animals in this initial population are inbred themselves. If the reference population chosen is too recent, or if there is a lot of missing information in pedigrees, then inbreeding coefficients may be severely underestimated.

Table 4.10. Calculating the inbreeding coefficient for animals resulting from matings between different types of close relative. In each case it is assumed that the common ancestor is not inbred itself, so the simpler version of the formula is used.

Type of mating	Number of individuals in pathway through common ancestor	Inbreeding coefficient (F)
Father - daughter	2 (B, A)	$F_C = (\frac{1}{2})^2$ $F_C = 0.25 \text{ or } 25\%$
		
Full brother - full sister	3 (C, A, D) + 3 (C, B, D)	$F_E = (\frac{1}{2})^3 + (\frac{1}{2})^3$ $F_E = 0.125 + 0.125$ $F_E = 0.25 \text{ or } 25\%$
		
Half-brother – half-sister	3 (B, A, C)	$F_D = (\frac{1}{2})^3$ $F_D = 0.125 \text{ or } 12.5\%$
		
Full cousins	5 (E, C, A, D, F) + 5 (E, C, B, D, F)	$F_G = (\frac{1}{2})^5 + (\frac{1}{2})^5$ $F_G = 0.03125 + 0.03125$ $F_G = 0.0625 \text{ or } 6.25\%$
		
Half uncle - niece	4 (B, A, C, D)	$F_E = (\frac{1}{2})^4$ $F_E = 0.0625 \text{ or } 6.25\%$
		
Half cousins	5 (D, B, A, C, E)	$F_F = (\frac{1}{2})^5$ $F_F = 0.03125 \text{ or } 3.125\%$
		

Matings between close relatives are relatively uncommon in farmed animal breeding today, because of the recognition of the high risk of undesirable side effects from high levels of inbreeding. However, high levels of inbreeding can occur also as a result of matings between less closely related animals for several generations. It is important to recognize this in designing breeding programmes. In this case the average inbreeding coefficient across a herd, flock or breed, and the rate at which this changes, are useful measures. Inbreeding coefficients can be calculated for large numbers of animals by a number of methods which are more amenable to computer programming than the method outlined above (Meuwissen and Luo, 1992; Mrode, 2014).

It is also useful to be able to predict the rate of inbreeding in advance, for example when comparing alternative designs for a breeding programme. Although it is relatively easy to predict rates of inbreeding for unselected populations, it is very difficult to do so accurately for populations under selection. Similarly, it is easier to predict rates of inbreeding when family sizes are equal than when they are not. Many of the simpler formulae underpredict rates of inbreeding, and so can give misleading results. There are more accurate formulae; however, these are too complex to describe here. (See Caballero (1994) and Villanueva *et al.* (1996) for more information on this problem.) The formulae below are reasonably simple, and illustrate some important points about rates of inbreeding, but they should be used with caution because they depend on assumptions which are rarely met in livestock breeding (equal family sizes, no selection, discrete generations).

The rate of inbreeding (usually abbreviated to ΔF) *per generation* can be calculated from (after Lush, 1945):

$$\Delta F \text{ per generation} = \frac{1}{8M} + \frac{1}{8F} \text{ (approx.)}$$

where:

M = the total number of males entering the population *each generation*; and

F = the total number of females entering the population *each generation*.

Since the number of females used is usually much larger than the number of males used in most livestock breeding programmes, it is the number of males that influences the rate of inbreeding most.

The rate of inbreeding *per annum* can be calculated from a modified version of this formula:

$$\Delta F \text{ per annum} = \frac{1}{8mL^2} + \frac{1}{8fL^2} \text{ (approx.)}$$

where:

m = the total number of males entering the population *each year*;

f = the total number of females entering the population *each year*; and

L^2 = the average generation interval in years, squared (i.e. multiplied by itself).

So, for example, in a closed flock of 100 sheep with 5 different rams and 35 new females entering the flock each year, and an average generation interval of 2 years:

$$m = 5$$

$$f = 35$$

$$L = 2$$

$$L^2 = 4$$

and

$$\Delta F = \frac{1}{8 \times 5 \times 4} + \frac{1}{8 \times 35 \times 4} = \frac{1}{160} + \frac{1}{1120}$$

$$= 0.00625 + 0.00089$$

= approximately 0.007 *per annum*, or 0.7%
per annum (of which 0.625% is due to males).

Table 4.11 shows the expected annual rates of inbreeding for several other breeding schemes, calculated using this formula. These results show that the number of parents selected *per annum*, and especially the number of sires selected, has a major impact on rates of inbreeding, as mentioned above. The values shown assume that selected animals are unrelated. However, in practice, related animals are often selected together because their breeding values are more alike than those of unrelated animals. This is particularly true when modern methods of predicting breeding values are used, which maximize the use of performance information from relatives. (These methods are explained in Chapter 7.) This co-selection of relatives increases the rate of inbreeding further. It is important in designing breeding programmes to achieve a balance between selecting fewer parents to achieve high selection intensities, which maximizes response, and using enough parents to keep inbreeding at acceptable levels. Traditionally, these two objectives have been pursued separately. However, the concept of optimizing genetic contributions of parents allows these goals to be pursued jointly (for a review, see Woolliams *et al.*, 2015). Statistical methods and software packages are available now to implement this approach (e.g. Wellmann, 2019).

Table 4.11. Predicted rates of inbreeding with different numbers of males and females selected. The generation intervals assumed are rounded versions of those shown in Table 4.5.

The results in Table 4.11 also show that shorter generation intervals increase the annual rate of inbreeding. This explains why turning over generations slowly is a useful method of minimizing inbreeding in conservation programmes, as explained in Chapter 3. Although there is little hard evidence available, it is generally accepted that absolute levels of inbreeding below 10%, or annual rates of inbreeding below 1% are unlikely to result in serious inbreeding depression. There is evidence that achieving a given level of inbreeding quickly through close matings is likely to be more harmful than achieving the same level of inbreeding more slowly through successive matings between less closely related animals – possibly because faster inbreeding allows fewer opportunities for deleterious alleles to be removed by natural selection (e.g. Pekkala *et al.*, 2014).

Summary

- It is important to be able to predict response to selection, in order to compare alternative breeding programmes, and so choose those which will achieve a high response to selection in a cost-effective way.
- When animals are selected on a single measurement of their own performance in one trait, the response to selection (R) per generation can be predicted by multiplying the selection differential (S) achieved by the heritability (h^2) of the trait selected on. The selection differential is the difference between the mean performance of selected animals and the overall mean of the group of animals from which they were selected. The heritability is the proportion of superiority of parents in a trait (i.e. the proportion of the selection differential) which, on average, is passed on to offspring. Alternatively, it can be defined as the additive genetic variation in that trait (or the variation in breeding values), expressed as a proportion of the total phenotypic variation in that trait. Heritabilities are expressed as proportions from 0 to 1, or as percentages from 0 to 100%.
- The ability to predict response to selection is important in planning breeding programmes and comparing alternative schemes. It is usually more useful to predict responses *per annum* rather than per generation. To do this, we have to divide the response per generation by the generation interval (the weighted average age of parents when their offspring are born; L) in the flock or herd concerned:
- From properties of the normal distribution, if we know what proportion of animals are selected as parents, we can predict their superiority in standard deviation units. This is called the standardized selection differential or selection intensity (i).
- The annual response to selection will be highest when the selection intensity is high, the heritability is high, and the generation interval is low. There is little that breeders can do about the heritability of a trait – this is largely a biological characteristic of the trait concerned. Within biological limits, breeders can increase selection intensities and decrease generation intervals.
- Records of performance from relatives are useful in selection. This is because related animals have genes in common, and so the performance of relatives can provide clues to the genetic merit or breeding value of the candidates for selection. Records of performance are often used from the animal's ancestors, the animal itself, the animal's full or half sibs, the animal's progeny, other more distant relatives of the animal, or combinations of these classes of relatives.
- To predict the response to selection when using information from relatives, a more general version of the formula is used. This involves the accuracy of selection (r) on any combination of records from the animal and its relatives. The accuracy is the correlation between animals' true breeding values for the trait(s) under selection, and the measurement(s) on which selection is based.
- The accuracy of selection on a single record of performance from a relative is calculated as the product of the proportion of genes they have in common and the square root of the heritability of the trait under selection. With records from greater numbers of relatives, in general: (i) the closer the relative the more valuable the record, (ii) initially, the more relatives of a given class, the higher the accuracy, although there are diminishing returns, (iii) initially, the higher the heritability, the higher the accuracy, and (iv) initially, the lower the heritability, the greater the proportional contribution which records from relatives make.

- Using records of performance from relatives can increase the accuracy of selection, but this benefit can be outweighed by increases in the generation interval. Usually records from ancestors and collateral relatives can be used to increase accuracy of selection without lengthening the generation interval. Using records from descendants usually increases the generation interval, and so there is a trade-off between increasing accuracy and increasing generation interval. Comparisons of different breeding schemes also need to consider the resources required, as well as predicted responses.
- Many traits of interest can be measured more than once over an animal's lifetime. Repeated records of performance can give additional clues to the breeding value of animals and enhance the accuracy of selection. The value of repeated records in selection depends on the repeatability. This is the correlation between repeated records from the same animal.
- The repeatability sets an upper limit to the heritability. Like heritabilities, repeatabilities range from 0 to 1, or from 0 to 100%. The higher the repeatability of a trait, the lower the value of repeated records in selection for that trait. As with the use of information from relatives, it is important to take into account any time lag in obtaining repeated records.
- Repeatabilities measure the similarity between repeated records of the same trait, assuming that performance measured in different years (or at other intervals) is controlled by exactly the same genes. However, there are often associations between different traits which are not likely to be influenced by exactly the same genes. The degree of association between these characteristics is measured by the correlation coefficient.
- Correlated response to selection can be predicted from the direct response if the genetic correlation between the two traits and the additive genetic standard deviation of the second trait are known.
- There are several limitations to the methods of predicting response outlined. The first is that selection itself causes a reduction in the amount of genetic variation in the first few generations of selection, so predictions of long-term responses to selection based on the initial variation in the population may be too high. However, there are more complex formulae to predict long-term response to selection. The second limitation is that after many generations of selection the variation in the trait of interest may become exhausted. But, predictions of response are usually satisfactory over the lifetime of most livestock breeding programmes. The final limitation is that the formulae presented predict the average response to selection. Individual breeding programmes may achieve higher or lower responses than this, by chance.
- Inbreeding is the mating of two related animals. Even when steps are taken to avoid it, inbreeding is inevitable in closed populations. There are two main reasons for wishing to limit inbreeding: (i) it reduces the amount of genetic variation in a population, and so can reduce response to selection; and (ii) it can lead to a decline in performance in traits associated with fitness, known as inbreeding depression. Effectively inbreeding and inbreeding depression are the opposite of crossbreeding and heterosis, respectively.
- The decline in genetic variation following inbreeding, and the amount of inbreeding depression, both depend on the amount of inbreeding. The closer the relationship between two animals which are mated, the greater the amount of inbreeding in the resulting offspring. Hence it is important to be able to measure the amount of inbreeding of existing animals, or the animals which could be produced from a particular mating. The amount of inbreeding is measured by the inbreeding

coefficient (abbreviated F). The inbreeding coefficient is formally defined as the probability that two alleles at any locus are 'identical by descent'.

- The inbreeding coefficient of an animal can be calculated from its pedigree by counting the number of individuals in each path through a common ancestor and back to the original animal. If there is more than one common ancestor, then this process is repeated for each one.
- It is also useful to be able to predict the rate of inbreeding in advance, e.g. when comparing alternative designs for a breeding programme. Many of the simpler formulae underpredict rates of inbreeding. However, the more accurate formulae are complex. The rate of inbreeding (ΔF) *per generation* can be calculated approximately from the total number of males and females entering the population *each generation*. The rate of inbreeding *per annum* can be calculated approximately from the total number of males and females entering the population *each year*.
- The number of parents selected *per annum*, and especially the number of sires selected, has a major impact on rates of inbreeding. It is important in designing breeding programmes to achieve a balance between selecting fewer parents to achieve high selection intensities, which maximize response, and using enough parents to keep inbreeding at acceptable levels. The co-selection of relatives increases the rate of inbreeding further. Also, shorter generation intervals increase the annual rate of inbreeding. Absolute levels of inbreeding below 10%, or annual rates of inbreeding below 1% are unlikely to result in serious inbreeding depression.

Appendix: The Connection Between the Two Formulae Presented to Predict Response to Selection

The original formula presented early in this chapter is:

$$R = \frac{i \times sd_p \times h^2}{L}$$

This can be converted to the second, more general formula in several steps. Firstly, we can expand the heritability to become:

$$h^2 = \frac{sd_A \times sd_A}{sd_p \times sd_p}$$

Substituting this in the formula above, instead of h^2 gives:

$$R = \frac{i \times sd_p \times sd_A \times sd_A}{L \times sd_p \times sd_p}$$

The sd_p on the top of the equation cancels out with one of the sd_p values on the bottom (i.e. both can be removed, to simplify the formula). Also, one of the sd_A values on the top of the equation, and

the remaining sd_p value on the bottom can be removed, and rewritten as h , giving (after Falconer and Mackay, 1996):

$$R = \frac{i \times h \times sd_A}{L}$$

This formula is appropriate if selection is on a single record of performance of the animal itself (i.e. $r = h$). For other types of selection h is replaced by r , the more general value of accuracy. If you do not follow the algebra involved in deriving these formulae, but want to prove that they really do mean the same thing, then re-calculate the response to selection for weaning weight in Simmental cattle, as in the earlier example, but using the new formula, with $h = 0.5$ (from $h^2 = 0.25$) and $sd_A = 17.5$ kg (from $h \times sd_p = 35 \times 0.5$).

References

- Animal Data Centre. (1997) UK Statistics for Genetic Evaluations (Dairy). February 1997. ADC, Rickmansworth, UK.
- Becker, W.A. (1984) Manual of Quantitative Genetics, 4th edn. Academic Enterprises, Pullman, Washington.
- Blasco, A., Sorensen, D. and Bidanel, J.P. (1998) Bayesian inference of genetic parameters and selection response for litter size components in pigs. *Genetics* 149, 301–306.
- Caballero, A. (1994) Developments in the prediction of effective population size. *Heredity* 73, 657–679.
- Dekkers, J.C.M. (1992) Asymptotic response to selection on best linear unbiased predictors of breeding values. *Animal Production* 54, 351–360. DOI: 10.1017/S0003356100020808.
- Falconer, D.S. and Mackay, T.F.C. (1996) Introduction to Quantitative Genetics, 4th edn. Longman, Harlow, UK.
- Fogarty, N.M. (1995) Genetic parameters for live weight, fat and muscle measurements, wool production and reproduction in sheep: a review. *Animal Breeding Abstracts* 63, 101–143. (Also, see erratum in Volume 63, No. 12.)
- Hayes, J.F., Cue, R.I. and Monardes, H.G. (1992) Estimates of repeatability of reproductive measures in Canadian Holsteins. *Journal of Dairy Science* 75, 1701–1706. DOI: 10.3168/jds.S0022-0302(92)77927-2.
- Lamberson, W.R. and Thomas, D.L. (1984) Effects of inbreeding in sheep: a review. *Animal Breeding Abstracts* 52, 287–297.
- Lush, J.L. (1945) Animal Breeding Plans, 3rd edn. Iowa State College Press, Ames, Iowa.
- MacNeil, M.D., Dearborn, D.D., Cundiff, L.V., Dinkel, C.A. and Gregory, K.E. (1989) Effects of inbreeding and heterosis in Hereford females on fertility, calf survival and preweaning growth. *Journal of Animal Science* 67, 895–901.
- Meuwissen, T.H.E. and Luo, Z. (1992) Computing inbreeding coefficients in large populations. *Genetics, Selection and Evolution* 24, 305–313. DOI: 10.1186/297-9686-24-4-30.
- Miglior, F., Burnside, E.B. and Dekkers, J.C.M. (1995a) Nonadditive genetic effects and inbreeding depression for somatic cell counts of Holstein cattle. *Journal of Dairy Science* 78, 1168–1173.
- Miglior, F., Burnside, E.B. and Kennedy, B.W. (1995b) Production traits of Holstein cattle: estimation of nonadditive genetic variance components and inbreeding depression. *Journal of Dairy Science* 78, 1174–1180.
- Morris, C.A., Baker, R.L., Hickey, S.M., Johnson, D.L., Cullen, N.G. and Wilson, J.A. (1993) Evidence of genotype by environment interaction for reproductive and maternal traits in beef cattle. *Animal Production* 56, 69–83. DOI: 10.1017/S0003356100006176.
- Mrode, R.A. (2014) Linear Models for the Prediction of Animal Breeding Values, 3rd edn. CAB International, Wallingford, UK.
- Nicholas, F.W. (1987) Veterinary Genetics. Oxford University Press, Oxford.
- Pekkala, N., Knott, K.E., Kotiaho, J.S., Nissinen, K. and Puurintien, M. (2014) The effect of inbreeding rate on fitness, inbreeding depression and heterosis over a range of inbreeding coefficients. *Evolutionary Applications* 7, 1107–1119. DOI: 10.1111/eva.12145.
- Saura, M., Almudena Fernández, A., Varona, L., Fernández, A.I., de Cara, M.A.R. et al. (2015) Detecting inbreeding depression for reproductive traits in Iberian pigs using genome-wide data. *Genetics Selection Evolution* 47, 1–9. DOI: 10.1186/s12711-014-0081-5. hal-01341281.
- Sewalem, A., Johansson, K., Williamson, M. and Lillpers, K. (1999) Inbreeding and inbreeding depression on reproduction and production traits of White Leghorn lines selected for egg production traits. *British Poultry Science* 40, 203–208. DOI: 10.1080/00071668.2014.989490.

- Simm, G., Bünger, L., Villanueva, B. and Hill, W.G. (2004) Limits to yield of farm species: genetic improvement of livestock. In: Sylvester-Bradley, R. and Wiseman, J. (eds) *Yields of Farmed Species. Constraints and Opportunities in the 21st Century*. Nottingham University Press, Nottingham, UK, pp. 123–141.
- Villanueva, B., Wray, N.R. and Thompson, R. (1993) Prediction of asymptotic rates of response from selection on multiple traits using univariate and multivariate best linear unbiased predictors. *Animal Production* 57, 1–13. DOI: 10.1017/S0003356100006541.
- Villanueva, B., Woolliams, J.A. and Gjerde, B. (1996) Optimum designs for breeding programmes under mass selection with an application in fish breeding. *Animal Science* 63, 563–576. DOI: 10.1017/S1357729800015459.
- Weller, J.I. (1994) *Economic Aspects of Animal Breeding*. Chapman and Hall, London.
- Wellmann, R. (2019) Optimum contribution selection for animal breeding and conservation: the R package optiSel. *BMC Bioinformatics*, 20, 25. DOI: 10.1186/s12859-018-2450-5.
- Wiener, G., Lee, G.J. and Woolliams, J.A. (1994a) Consequences of inbreeding for financial returns from sheep. *Animal Production* 59, 245–249. DOI: 10.1017/S000335610000773X.
- Wiener, G., Lee, G.J. and Woolliams, J.A. (1994b) Effects of breed, rapid inbreeding, crossbreeding and environmental factors on fleece weight and fleece shedding in sheep. *Animal Production* 59, 61–70. DOI: 10.1017/S0003356100007510.
- Wolc, A., Arango, J., Settar, P., O'Sullivan, N.P. and Dekkers, J.C.M. (2017) Repeatability vs. multiple-trait models to evaluate shell dynamic stiffness for layer chickens. *Journal of Animal Science* 95, 9–15. DOI: 10.2527/jas.2016.0618.
- Woolliams, J.A., Berg, P., Dagnachew, B.S. and Meuwissen, T.H.E. (2015) Genetic contributions and their optimization. *Journal of Animal Breeding and Genetics* 132, 89–99. DOI: 10.1111/jbg.12148.
- Wray, N.R. and Simm, G. (1991) BLUP for Pedigree Beef and Sheep Breeders. SAC Technical Note T265. The Scottish Agricultural College, Perth, UK

3.1.3

Simm et al. (2020)

Page 81-87, and 104-107

5. Tools and Technologies in Animal Breeding

Introduction

Genetic improvement is one of the most effective strategies available for enhancing the performance of farmed animals. Some approaches, such as within-breed selection, are relatively slow compared to some other methods, such as improved feeding, but it is permanent and cumulative, and in most cases, it is highly cost effective and sustainable. Populations of animals of high genetic merit are needed to achieve high efficiency, competitiveness, and both economic and environmental sustainability, in any livestock industry. Genetic improvement has generally been used very effectively in the pig, poultry and dairy industries of many wealthier countries, and increasingly so in certain aquaculture sectors. However, it has been used less effectively until recently in other ruminant sectors. The use of objective genetic improvement of farmed animals has been much less widespread in low- and middle-income countries to date, and this represents a major opportunity for improving productivity and livelihoods. The aim of this chapter is to discuss some of the tools and technologies which are being used, or could be used, to lead to more effective genetic improvement programmes in farmed animals. While most of these tools and technologies are relevant in all species, they are likely to be applied in a species-specific manner. Some such applications are described in later chapters.

Rates of genetic improvement depend on four main factors, as outlined in Chapter 4: (i) the selection intensity achieved; (ii) the accuracy with which genetic merit in the trait of interest is predicted; (iii) the amount of genetic variation in the trait of interest; and (iv) the generation interval. Generally speaking, the greater the selection intensity, accuracy and genetic variation, and the shorter the generation interval, the greater the annual rate of genetic improvement. The main opportunities for breeders to accelerate rates of improvement are through choice of the most accurate methods of predicting breeding values, and by maintaining high selection intensities and short generation intervals. However, there are often biological limits on the extent to which selection intensity and generation interval can be altered. As a result of their earlier sexual maturity, and their higher reproductive rates, it is possible to achieve greater selection intensities and shorter generation intervals in pigs, poultry and some aquaculture species than in ruminants. Largely as a result of this, annual rates of genetic improvement in pigs, fish and poultry are often up to double those predicted or achieved in ruminants (Smith, 1984; Gjedrem and Rye, 2018).

The scope for increasing rates of genetic gain through the wider use of BLUP and genomic methods of predicting breeding values is discussed in later chapters. Rates of genetic gain in overall economic merit will be maximized by ensuring that breeding goals include all heritable traits which are of major economic importance, and include none that are of very minor importance, or are not heritable. Similarly, selection indexes should include all the available measurements that make a significant contribution to predicting merit in breeding goal traits. There is considerable scope for the use of more comprehensive breeding goals and criteria in most livestock breeding programmes. For

instance, including traits affecting longevity, health and greenhouse gas emissions in dairy cattle breeding indexes is likely to make an increasingly important contribution to overall economic progress, cow welfare, and the sustainability of breeding programmes. Also, the inclusion of traits related to leg health and reduced mortality in addition to growth rate in broiler birds could improve sustainability of the breeding programme and increase public acceptability. Similarly, a better understanding of the relationships between production traits and traits conferring adaptation to harsh environments could help in the production of more sustainable breeding programmes for hill or tropical sheep and beef cattle in these environments.

In most cases the tools to achieve these improvements are already available but they need to be finetuned and applied more widely. For instance, the methodology for producing more comprehensive breeding goals and indexes is well developed. However, indexes need to be tailored to the particular goal, by obtaining estimates of the relevant *genetic parameters* and *economic values*. Although the methodology is available, this is by no means a trivial task. In other cases, new technologies can contribute to accelerating genetic improvement. For example, new scanning techniques can improve rates of progress in carcass characteristics, by providing more accurate predictions of the carcass composition of candidates for selection. New techniques for automatic identification of animals and data capture could improve the accuracy of selection for a range of performance traits in many livestock species; these are especially useful in difficult-to-measure traits such as feed intake and disease incidence. Similar techniques could provide vast amounts of additional information on the carcass weights and grades of commercial meat animals. This could be of considerable value in increasing the accuracy of evaluation of related purebred animals. Methods to estimate greenhouse gas emissions from individual ruminant animals will contribute greatly to selection programmes seeking to incorporate reduction of emissions in their breeding goals.

In addition to data capture tools, the other types of new technologies which can have a major impact on rates of genetic improvement include reproductive and genomic technologies. The current and potential future reproductive, genetic and data capture technologies, together with their possible impact on genetic improvement of livestock, are discussed in detail in the next three sections. Any new technique needs to have a favourable cost–benefit ratio to be widely adopted. These issues are not discussed in detail here, partly because many of the techniques are at an early stage of development, and so both success rates and costs are difficult to estimate. However, it is worth making two general comments about the cost–benefits of new technologies. Firstly, it will usually be easier to justify the use of more expensive technologies in order to accelerate genetic improvement programmes in elite populations, than to justify their use for dissemination of improvement to commercial tiers. This is because of the higher average value of animals produced from elite populations. Secondly, it will be easier to justify the use of more expensive technologies in species with a higher average *per capita* value of elite and commercial animals or those species where reproductive rates are high, such as pigs, poultry and aquaculture species.

In many industrialized countries over the last few years there has been growing public interest in methods of food production, and particularly in the welfare of farmed animals. Particular concerns are expressed about the potential animal welfare or ethical implications of some of the reproductive and genetic technologies outlined in this chapter, and these issues are discussed in Chapter 14.

Reproductive Technologies

It is possible to achieve much higher selection intensities in species or breeds with a high reproductive rate than in those with lower reproductive rates. Similarly, shorter generation intervals can be achieved in those species or breeds which reach sexual maturity at a younger age. It is largely because of biological advantages in these reproductive characteristics that higher rates of genetic change are possible in fish, poultry and pigs than in ruminants. In dairy cattle the main traits of interest are sex-limited and measured fairly late in life, which compounds the disadvantage that cattle suffer in reproductive rate. There are several reproductive technologies which can accelerate progress in genetic improvement programmes in livestock. These include artificial insemination (AI), multiple ovulation and embryo transfer (MOET), *in vitro* production of embryos (i.e. production by laboratory culture), sexing of semen or embryos, and *cloning* (i.e. mass production of identical embryos). These techniques are outlined briefly below, and their potential impacts are discussed. More details on the techniques themselves are given in reviews or books by Woolliams and Wilmut (1989) Wilmut *et al.* (1992), Robinson and McEvoy (1993), Gordon (1994, 1996, 1997), Luo *et al.* (1994), Hernandez Gifford and Gifford (2013), Rutten *et al.* (2013) and Niemann and Wrenzycki (2018).

In addition to their potential value in accelerating response to selection in breeding programmes, many of these techniques also have the potential to accelerate dissemination of genetic improvement from the elite to the commercial tiers of livestock industries. Techniques of value in selection will generally be useful in dissemination but their cost and ease of use may limit their role in dissemination. However, not all techniques of value in dissemination will be useful in selection programmes. So, the value of each of the techniques in dissemination is also discussed.

Artificial insemination

Artificial insemination has been available to cattle breeders for many years, to enhance the reproductive rate of males (see Foote (2002) for a review of AI development). The development of reliable techniques for extending (i.e. diluting to allow wider use) and freezing cattle semen has augmented this benefit. AI allows much higher selection intensities amongst males than those possible with natural mating. Also, the desired number of progeny can often be produced sooner by AI than by natural mating, so male generation intervals can be reduced. AI can contribute to more accurate evaluation of genetic merit as well, by permitting large-scale progeny testing in many herds or flocks. As a result, substantial rates of genetic improvement have been achieved in many countries through the use of AI in well-designed dairy cattle breeding schemes. Progress has been particularly high when the use of AI has been coupled with accurate techniques for predicting breeding value, as outlined in Chapter 7.

Although the use of AI is less widespread in beef cattle breeding than in dairy cattle breeding, the technique can have a similar impact. In many countries, one of the major contributions of AI to beef breeding programmes is to create *genetic links* between herds, which allow across-herd genetic evaluations.

The fact that AI in cattle is relatively cheap and simple, and often allows access to very reliably proven, high genetic merit animals means that it is currently the most effective method of dissemination of genetic improvement to commercial herds. This is particularly true in dairying, where

commercial herds rely heavily on AI, and have as good access as elite breeders to high genetic merit bulls. Most commercial beef cows are kept in more extensive production systems than dairy cows. This makes oestrous detection more difficult and hence limits the use of AI for dissemination. However, there is growing use of oestrus synchronization in commercial beef herds to make AI more practical, and so allow access to bulls of higher merit.

To date, AI has had a much smaller impact in most sheep breeding programmes than in cattle breeding programmes. This is largely because of the poor conception rates which usually accompany the cervical insemination of frozen semen. However, the technique has had an important impact in schemes where the use of fresh semen is practical, including dairy sheep breeding programmes, or in schemes involving the laparoscopic intra-uterine insemination of frozen semen. (With laparoscopic AI, the uterus is viewed, and semen is inserted into the uterus via small incisions in the abdomen.) For instance, the strategic use of laparoscopic AI, and in some cases cervical AI, to create genetic links across flocks, has been central to the success of sheep sire referencing schemes, as outlined in Chapter 10. There is no doubt that new techniques which produce high conception rates, and use less invasive insemination methods than laparoscopy, could have a major impact on both the rates of genetic gain and the dissemination of genetic improvement in the sheep industries of many countries. The use of objective selection in sheep breeding programmes and the evidence of high rates of improvement achieved in several large co-operative breeding schemes, make the development and application of improved techniques for dissemination particularly timely.

In lower income countries, especially in sub-Saharan Africa, AI has been of limited use, even in dairy cattle breeding. Some of the major challenges include lack of reliable plants to produce the liquid nitrogen needed for freezing semen, high cost, lack of trained AI technicians, and small farms that are widely dispersed and often inaccessible. In a review of world statistics on AI in cattle, Thibier and Wagner (2002) reported 646 bulls in semen collection centres in Africa compared to 9627 bulls in North America and 20,785 in Europe. In an attempt to address some of the constraints, the Bill and Melinda Gates Foundation are currently implementing a project called PAID (Public Private Partnership for AI Delivery, 2019) in collaboration with private AI companies to increase the efficiency of AI usage in several African countries.

Multiple ovulation and embryo transfer

Over the last few decades, increasingly reliable procedures have been developed for superovulation, embryo recovery, embryo freezing and embryo transfer in cattle (Woolliams and Wilmut, 1989). Similar procedures have been developed for sheep but, unlike the situation in cattle, embryo recovery and transfer techniques are only practicable at present with laparoscopy or surgery (Dingwall and McKelvey, 1993; Gordon, 1997). A number of applications of embryo procedures have been proposed or practised in cattle and sheep breeding. These include uses in: (i) within-breed genetic improvement programmes; (ii) the international trading of genetic material (offering potential advantages in economy, animal welfare and disease control); (iii) accelerating breed substitution by multiplication of newly introduced breeds; and (iv) conservation of genetic material by freezing embryos from valuable individual animals, or from rare or endangered breeds or species. Additionally, several of the new reproductive or genetic procedures discussed in this chapter hinge on the use of embryo transfer.

MOET potentially offers similar benefits in selection of females to those offered by AI in males (though in practice the benefits are often smaller). That is, female selection intensities can be increased, female generation intervals can be reduced, and the accuracy of evaluating embryo donors, or full sibs created by MOET, can be increased. Since the mid-1970s there have been several studies on the potential impact of MOET on the rates of genetic improvement in ruminants. The earliest of these studies predicted that rates of improvement in beef cattle, dairy cattle and sheep could be increased up to twofold compared to conventional breeding schemes (Land and Hill, 1975; Nicholas, 1980; Nicholas and Smith, 1983; Smith, 1986).

As a result of the early work, commercial dairy cattle breeding schemes based on MOET were initiated in several countries in the mid- to late 1980s. Experimental breeding schemes involving MOET were also established in beef cattle and sheep at about the same time. However, further theoretical work from the mid-1980s onwards showed that the initial results were very sensitive to alterations in some key assumptions. For example, most of the early studies used optimistic success rates for MOET, and these have been difficult to achieve on a 'field scale' in practice. Also, the early studies usually ignored the fact that genetic variation, and hence response to selection, is reduced by the process of selection, and by inbreeding – the latter being a usual consequence of effective selection in domestic animals. These factors appear to be particularly important in small, closed nucleus populations such as those in which the use of MOET was first envisaged. The large variability in numbers of transferable embryos recovered per flush was also ignored in many early studies. This variability is expected to have an unfavourable impact on rates of genetic gain and, especially, on rates of inbreeding, since fewer animals make a bigger contribution to the genetic makeup of the next generation. Advances in the theory of predicting rates of inbreeding also showed that these had been significantly underestimated in most of the early studies (Villanueva and Simm, 1994; Nicholas, 1996).

Much effort has been directed at the design of breeding schemes to overcome these high rates of inbreeding, and several successful techniques have been identified (Toro *et al.*, 1988; Woolliams, 1989; Grundy *et al.*, 1994; Villanueva and Simm, 1994; Wray and Goddard, 1994a; Caballero *et al.*, 1996; Meuwissen, 1997). These include:

- the use of factorial mating designs (mating each donor female to different selected males in successive matings, rather than to the same male);
- using selected parents only once;
- equalizing family sizes (this will be easier to achieve with new techniques producing higher yields of embryos);
- using more than one male from each selected full sib family;
- deliberately choosing less closely related animals; and reducing the emphasis on ancestors' performance when calculating BLUP EBVs (because BLUP methods use records of relatives they tend to lead to selection of more closely related animals, and so to higher rates of inbreeding; this effect can be reduced by altering the emphasis on records from relatives); and
- minimizing the average coancestry (or half the average *additive genetic relationship*) among selected parents or maximizing genetic improvement while restricting the average coancestry among selected parents.

Use of some of these techniques, either alone or in combination, is expected to reduce inbreeding substantially, with little or no change in genetic gain. So, with appropriate modifications to design, MOET schemes can probably offer increased rates of response compared to conventional schemes. However, these rates of response are still unlikely to be as high as first predicted. For example, in beef MOET schemes, rates of gain may be 30% higher than in conventional schemes, at acceptable levels of inbreeding, rather than 100% higher, as first predicted (Villanueva *et al.*, 1995). In future, the advantage to MOET schemes may be augmented as a result of the use of techniques which result in improved yields of transferable embryos, or through the use of some of the new reproductive or genetic technologies outlined below. For example, Pryce and Daetwyler (2012) and Granleese *et al.* (2015) examined the incorporation of genomic selection into MOET schemes by using or reviewing simulation studies. The latter authors estimated that the use of genomic selection with MOET schemes should result in increases in the rates of genetic gain of between 38% and 76%, depending on the level of inbreeding modelled, when compared to using AI or natural mating.

There have been few recent studies of the benefits of MOET in sheep. However, for meat breeds, it is probably reasonable to expect similar benefits to those in beef MOET schemes. In wool breeds, the benefits appear to be similar to those outlined for beef cattle too (Wray and Goddard, 1994b; Bergstein-Galan *et al.*, 2019).

In theory, MOET could be a highly effective technique for the dissemination of genetic improvement to commercial sectors of the livestock industries. Since embryos have already passed the hurdle of fertilization, they have the potential to produce higher calving or lambing rates than AI, though this potential is still far from being realized. Similarly, embryos already contain the full complement of chromosomes necessary for their development. So, MOET can deliver populations of commercial animals with 100% of their genes from elite sector parents, whereas AI used on commercial females produces animals with only 50% of their genes from the elite sector. However, with current MOET techniques, each donor female produces relatively few embryos, at least compared to the number of doses of semen which can be produced by each male. Partly for this reason, but also because of the more complex techniques involved, MOET remains far more expensive than AI, and so less attractive as a means of dissemination. Some newer techniques that have the potential to allow the wider use of embryo transfer are discussed next.,

In vitro production of embryos

Over the last few decades, a great deal of effort has gone into developing techniques for the *in vitro* maturation, fertilization and culture of eggs from farm animals, as well as humans. *In vitro* literally means in glass, as opposed to in the body. The rationale for this research in humans is to allow certain types of infertility to be overcome (e.g. that due to blocked Fallopian tubes), or to allow the use of donor semen or eggs in cases of complete infertility of one partner. The main aim of developing these techniques in farm animals is to allow the use of the thousands of eggs present in the ovaries of female animals at birth, most of which never develop to the point of ovulation (Betteridge *et al.*, 1989; Gordon and Lu, 1990; Wilmut *et al.*, 1992; Gordon, 1994).

One of the earliest intended uses of *in vitro*-produced embryos was to improve the beef merit of calves from dairy or suckler cows, by creating a supply of embryos with 3/4 or 7/8 beef genotype.

Initially, the main source of eggs was the ovaries of slaughtered beef heifers. Companies were established in several countries, including Britain and Ireland, to collect eggs from beef heifers with a high proportion of continental beef breeds in their genetic makeup, and to produce embryos from these by maturing them and then fertilizing them with semen from high merit, proven bulls. These embryos were then marketed for transfer into beef suckler cows or dairy cows. Transfers were made either singly or to create twins by transferring an *in vitro*-produced embryo into cows already carrying a natural embryo or by transferring two *in vitro*-produced embryos. Despite a ready supply of ovaries from slaughtered heifers, early techniques produced few transferable embryos per ovary. Also, some *in vitro* culture techniques are implicated in the birth of very large calves, generally with associated calving difficulties (Kruip and den Daas, 1997). Partly for these reasons, some of the original companies are no longer in business.

All purebred cows are slaughtered or die of natural causes at some stage, but most are no longer prime candidates for selection by the time this happens, so recovering eggs *post mortem* is not so beneficial. Hence, the method outlined above using slaughtered heifers is particularly suitable for the dissemination of genetic improvement. The majority of donors are likely to be crossbred animals of unknown genetic merit, with high performance in commercial beef traits. Despite the lack of predictions of genetic merit of donors, selection on the basis of breed or crossbred type, or crude selection on phenotype, should be sufficient to ensure that their merit for beef production is better than that of most dairy cows, and many suckler cows. Also, the technique requires semen from only a small number of very highly selected bulls to fertilize eggs, so it allows very high male selection intensities and accuracies to be achieved. This will usually compensate for possible deficiencies in the selection procedures for donor females. With improvements in the techniques involved, and selection of appropriate parents, the *in vitro* production of embryos could improve both quality and uniformity of beef production in future.

Techniques have been developed to allow the recovery of unfertilized eggs directly from the ovaries of live cows. These involve collection of eggs through an ultrasonically guided needle inserted into the ovary, usually via the vagina (Kruip, 1994; Boni, 2012). This type of recovery is called *in vivo* aspiration of oocytes, or *ovum pick up* (OPU). It has several potential advantages compared to recovery of eggs from slaughtered cows, or to conventional embryo recovery techniques:

- Purebred animals of high genetic merit can be used as donors, so the technique is of potential benefit in genetic improvement and not just in dissemination.
- The technique can be applied to produce embryos in a more planned manner than with *post-mortem* recovery.
- It is possible to collect oocytes from younger donors in this way than with conventional embryo recovery techniques.
- It is possible to collect oocytes from donors in the early stages of pregnancy.
- Eggs can be collected from donors on a weekly basis, allowing tens or potentially hundreds of embryos to be produced from the same donor.

Because it is an invasive technique it does require skilled veterinary input. However, there appears to be no evidence of injury to the donor, nor any adverse effects on subsequent reproductive performance.

OPU is being used in several countries to produce a higher number of transferable embryos than that produced by conventional MOET. Increasing the yield of transferable embryos in this way could have a major impact on the rate of gain achievable in MOET breeding schemes, possibly allowing rates of gain up to 34% above those possible in conventional progeny testing schemes (Lohuis, 1993). Also, high yields of *in-vitro*-produced embryos make some of the methods of controlling inbreeding more practical (e.g. equalizing family sizes, use of factorial mating designs). Embryos produced by OPU are likely to be used in genetic improvement programmes in elite herds, rather than for mass dissemination of improvement to the commercial sector, unless the yield of transferable embryos increases dramatically, or it is combined with other new reproductive technologies such as embryo cloning.

In theory, *in-vitro*-produced sheep embryos would have many of the advantages already mentioned for cattle embryos. However, the techniques are less well developed for sheep, and the lower value of the end product is likely to limit applications in sheep breeding and especially in commercial sheep production.

Semen and embryo sexing

Sexing semen has been a major aim of reproductive technologists for decades. A major breakthrough came with the publication of the paper by Johnson *et al.* (1989) reporting success with rabbits. Most of the methods tested have used real or assumed physical differences between sperm bearing X and Y chromosomes as the basis for separation. For example, separation methods aimed at exploiting differences in mass, surface charge, antigenic properties and buoyant density of X- and Y-bearing sperm have been investigated, but success rates and repeatability have usually been low (White, 1989; Cran and Johnson, 1996). In the last few years, a more reliable technique for *semen sexing* has been developed. The technique uses the fact that sperm bearing the X and Y chromosomes differ slightly in their DNA content. Using a technique called flow cytometry, it is possible to detect these differences in size, and to sort sperm into two groups accordingly. The technique involves staining sperm with a fluorescent stain (e.g. a DNA-staining solution such as Hoechst 33342) and using differences in fluorescence when the cells pass through a laser beam to identify and sort X- and Y-bearing sperm. A number of reviews of this technology are available (Seidel and Garner, 2002; Seidel, 2012; de Graaf *et al.*, 2014). Currently, the method of sexing mammalian sperm using a flow cytometer and measuring DNA content of sperm through the fluorescence of the DNA-bound Hoechst 33342 is the only commercially viable method to sex-sort mammalian sperm and obtain pregnancies. Recently, the process has undergone several improvements both in sperm handling and preparation for sorting. Also, significant enhancements in sorter technology such as advanced digital processing, multiple sensor heads and automation has made this process more efficient and conception rates now compare favourably with conventional semen (Sharpe and Evans, 2009; Evans, 2010; Vishwanath and Moreno, 2018). This methodology is now used in several livestock species, including dairy and beef cattle, pigs, sheep, goats, deer and horses.

Several approaches have been taken to develop techniques for sexing embryos (van Vliet *et al.*, 1989; White, 1989; Woolliams and Wilmut, 1989). These are based on removing a small number of cells from embryos at the 16-cell stage or later stages of development. In some cases, the sex of the embryo can be determined by direct observation of a preparation of chromosomes using a

microscope (karyotyping), to confirm the presence of either two X chromosomes or one Y and one X chromosome. Alternative approaches include the use of an immunological test for the H-Y antigen produced only by male embryos, using sex-linked differences in enzyme activity, or using some of the molecular techniques described later in this chapter to detect sequences of DNA specific to the Y chromosome (Zoheira and Allam, 2010; Kageyama and Hirayama, 2012). Perhaps the sector most in need of embryo sexing is the egg-laying chicken industry where the production of many unwanted males continues to be a welfare problem. (Unlike mammals, in birds females are the heterogametic sex so sexing semen is not relevant.) Invasive methods based on identifying sections of the sex chromosomes taken from biopsies of the eggs look promising to solve this long-term problem (He *et al.*, 2019).

In genetic improvement programmes, sexing of either semen or embryos could be useful to increase the selection intensity applied to females by producing more of them. This appears to be of little value when performance records are available on both sexes, and where the total number of animals tested is fixed, as an increase in the number of females implies a reduction in the number of males, and therefore a reduction in male selection intensities. However, when performance recording is sex-limited, as with milk production, there may be benefits in creating more animals of the recorded sex. This may be especially beneficial in nucleus schemes, if they already have access to high genetic merit males which have been accurately tested in a separate population (Colleau, 1991; Villanueva and Simm, 1994).

The use of relatively cheap, reliable techniques for sexing semen in large enough quantities for widespread conventional AI should lead to major improvements in the dissemination of genetic improvement and in the efficiency of animal production. In meat production systems there is interest in using 'female' semen to breed replacement animals from the highest merit females in the herd or flock. Since males usually show higher growth rates and leaner carcasses, there may be a preference for 'male' semen to produce animals for slaughter. However, so-called once-bred heifer systems, using only 'female' semen, may produce the highest overall efficiency (Taylor *et al.*, 1985). In these systems, each heifer leaves a replacement female calf before being slaughtered for beef production. In conventional dairy cattle systems, 'female' semen is used to breed replacements from the best heifers or cows, releasing a higher proportion of the herd for mating to 'male' semen from a beef breed, or for crossing to a beef bull. This is happening increasingly in the dairy sector in many countries.

Similarly, the development of techniques for the production of large numbers of cheap, sexed embryos could lead to the use of high merit female embryos to produce replacements, and cloned male embryos to produce slaughter animals, or the development of once-bred heifer systems based only on female embryos. In dairy herds, elite female embryos could be used to produce replacements, and male embryos of high beef merit and a proven record for calving ease, could be used in the majority of cows. Also, improved techniques for embryo multiplication could increase the benefits and practicality of crossbreeding. In cattle and sheep, crossbreeding is sometimes impractical because the low reproductive rate of these species means that many purebreds are needed to produce a regular supply of F₁ females. In some cases, there are too few breeds with high genetic merit in the traits of interest to sustain a rotational crossing scheme. However, improved reproductive techniques could allow a plentiful supply of F₁ (or other) embryos to produce replacement females, from a relatively small population of elite animals of the component pure breeds. This could lead to replacement policies in cattle and sheep similar to those already employed

by conventional means in the commercial tier of the pig and poultry industries, where most replacement females are purchased crossbreds. So far this has had little impact on any livestock industry but the newer DNA-based methodologies for embryo sexing look promising for both cattle and chickens.

Embryo cloning

The production of groups of identical embryos from a single original embryo is often termed embryo cloning, or embryo multiplication. This can be achieved in one of two ways at present. The first involves physically splitting embryos into two halves (or four quarters). If the original embryo was at an early stage of development, the resulting halves need to be inserted into an empty *zona pellucida* or 'egg shell', before transfer. If the embryo is at a later stage of development, the halves can be transferred without this procedure (Woolliams and Wilmut, 1989). Splitting into more than four parts is not successful, as there are too few cells for normal development. Also, repeated bisection of embryos after further culture is not successful because it leads to too great a disparity between the physical composition and true chronological stage of development of the embryos. Embryo bisection can be employed to increase the yield of transferable embryos, particularly when donors are rare or very valuable, or in MOET *nucleus breeding schemes* where a target number of embryos is needed from each donor.

A larger number of identical embryos can be produced by the procedure of *nuclear transfer*. This involves removing the *zona pellucida* from a 16-, 32- or 64-cell embryo, and separating the identical cells within. Each of the resulting cells can be inserted into an unfertilized egg from which the nucleus containing the single copy of chromosomes has been removed. An electric current causes fusion of the introduced embryonic cell and the unfertilized egg, and the new embryo then develops as if newly fertilized (Woolliams and Wilmut, 1989). The advantage of this technique is that it produces larger numbers of identical animals, although the techniques involved are more complex. Initially, nuclear transfer was only accomplished using early embryos, or cells from early embryos. However, in 1996 it was reported that viable cloned embryos, and subsequently live cloned lambs, had been produced by transferring cells which were originally derived from embryos, but had been cultured and multiplied in the laboratory (Campbell *et al.*, 1996; Campbell and Wilmut, 1997). The ability to multiply embryonic cells in the laboratory, and achieve successful development of the resulting embryos following nuclear transfer, obviously gives the potential for far greater numbers of identical animals to be produced from a single pair of elite parents. It also creates potential opportunities for the rapid multiplication of embryos which have been genetically modified in some way (e.g. by *gene transfer* or editing).

In 1997 it was reported that a viable lamb ('Dolly') had been produced following nuclear transfer from a cultured cell originating in an adult ewe – the first time that a mammal has been produced from a cell derived from an adult (Fig. 5.1; Wilmut *et al.*, 1997). This presents new opportunities for adult cloning, since clones can be derived from a single animal of proven performance, rather than from embryos produced by two proven parents. By the year 2007, 10 years after Dolly's birth, an additional 18 mammalian species including pigs, cattle, goats and horses had been cloned from adult

Fig. 5.1. Cloned sheep. The two sheep on the left (*Megan* and *Morag*) were produced by nuclear transfer of cultured cells originally derived from an embryo. The sheep on the right (*Dolly*) was produced by nuclear transfer of a cultured cell originating from an adult ewe – the first time that a mammal had been produced from a cell derived from an adult. (Courtesy of Professor Sir Ian Wilmut.)



cells using somatic cell nuclear transfer (Baguisi *et al.*, 1999; Wells *et al.*, 1999; Polejaeva *et al.*, 2000; Galli *et al.*, 2003). The ability to create clones created great interest in agriculture and the biomedical industry, and Dolly's birth in particular, stimulated both commercial and wider public interest in cloning.

In genetic improvement programmes, cloning has the potential to be used to produce many animals of the same genotype in order to improve the accuracy of evaluation, or to allow evaluation of traits normally measured post slaughter on some members of the cloned group. This would involve implanting some embryos from each cloned line to produce animals for testing, and freezing others to allow subsequent use (or further cloning) of the best tested cloned lines in breeding or dissemination programmes. Cloning could theoretically be of value in breeding programmes under certain circumstances, for example in dairy cattle nucleus schemes, if there are elite males available which are tested in a separate population. One potential advantage of the use of cloning would be better utilization of non-additive genetic variation. Conventionally, selection for quantitative traits within breeds attempts to identify animals with the highest additive genetic merit. It is difficult to make use of non-additive genetic variation in selection since particular combinations of non-additive genes which lead to the high merit of parents get split up and mixed between one generation and the next. However, comparison of cloned lines created from different parents would allow selection of those lines with the best combination of additive and non-additive genetic merit. Since the animals produced from the same cloned line have identical genotypes, the favourable genotype could be recreated exactly by further cloning. Similarly, cloned lines could be produced from crosses between breeds to exploit heterosis if this is important for the traits of interest. Clones with the optimum proportion of each parent breed could be produced, so minimizing the problems seen with rotational crossing, which inevitably produces many animals with sub-optimal proportions of the different breeds involved. However, if cloning is considered only in the context of closed breeding schemes, with fixed numbers of animals tested, then the expected benefits generally diminish or disappear, as keeping more identical animals means that fewer different families can be kept, and so selection intensities will be reduced (de Boer *et al.*, 1994; Villanueva and Simm, 1994).

While the benefits of cloning in genetic improvement may be limited, the potential of the technique to accelerate dissemination of genetic improvement to commercial herds or flocks is great. However, a number of drawbacks have prevented this becoming a commercial reality to date. Firstly, fully reliable, cost effective and scalable methods for cloning and delivery are not yet available. Further,

a disadvantage of the technique is the risk of producing very large numbers of animals which subsequently turn out to be susceptible to disease or to be unsuitable to new production methods. However, it should be possible to reduce such risks by maintaining reasonable genetic diversity in the conventionally bred elite populations and producing several or many separate cloned lines from these.

The future use of cloning may be in combination with other emerging technologies, such as genomic selection and gene editing, as discussed further below (Watson, 2018). For example, Carlson *et al.* (2016) combined gene editing with cloning to substitute the polled allele from Angus cattle into a Holstein cell line. Gene editing was used to introduce the Angus allele into a Holstein cell line and those cells in which the desired allele was present were selected for cloning. This approach significantly reduced the time it would take to introduce the trait through conventional breeding alone.

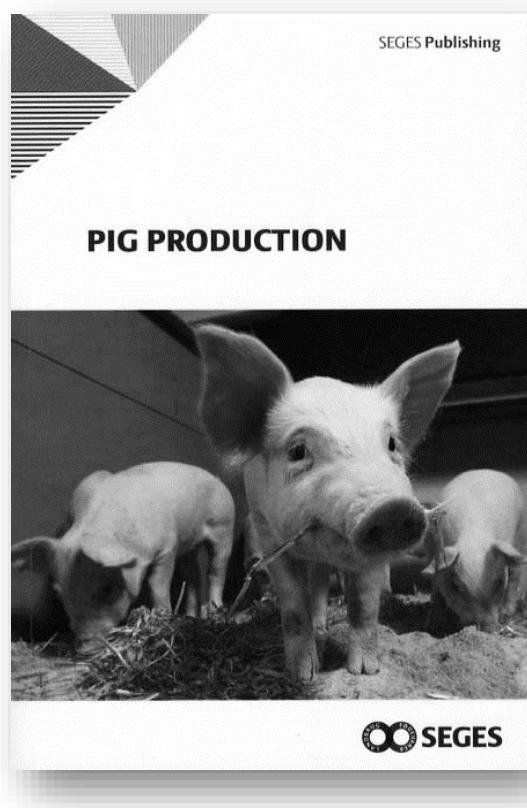
Kasinathan *et al.* (2015) reported the use of several reproductive technologies, cloning and genomic selection, in combination, to reduce the generation interval in cattle. They treated Jersey dairy cows to stimulate oocyte production. Oocytes were collected and used to produce IVF embryos. Multiple IVF embryos were transferred into surrogate mothers and allowed to develop for 21 days then flushed and collected. Cell lines were generated from the collected material and sent for DNA analysis to identify those of high genetic merit using genomic selection. The high genetic merit cell lines were used to produce cloned calves using nuclear transfer.

Summary

- New reproductive and molecular genetic technologies could lead to more effective genetic improvement programmes in livestock, or to more rapid dissemination of improved genes from elite to commercial sectors of the livestock industries. While most of these technologies are relevant in all species, some are likely to be of particular value in ruminants.
- Artificial insemination is already widely used in cattle breeding but, largely because of technical difficulties, it is not yet so widely used in sheep breeding. AI can allow higher male selection intensities, shorter male generation intervals and higher accuracy of selection, as well as providing genetic links across herds or flocks. Hence, it is a highly effective method for increasing rates of genetic improvement. It is also an extremely useful method for dissemination.
- Multiple ovulation and embryo transfer (MOET) potentially offers similar benefits in the selection of females to those offered by AI in males. However, in practice the benefits are often smaller. As a result of this and its relatively high cost, the technique is largely confined to specialized breed improvement schemes (often involving elite nucleus populations), to facilitate international trade in genetic material, and to accelerate multiplication of newly introduced breeds.
- *In vitro* production of embryos can dramatically increase embryo yield compared to conventional MOET. This could accelerate rates of improvement in breeding schemes, especially when oocytes are collected from live elite donors (ovum pick up). However, it could also allow more effective dissemination to the commercial sector, for example when oocytes are collected from slaughtered beef heifers of high performance, for transfer to dairy or suckler cows of lower beef merit.
- While the use of reproductive technologies such as MOET can substantially increase rates of genetic improvement, rates of inbreeding often increase proportionately more. However, much of this gain can often be achieved with minimal increase in inbreeding by modifications to the design of the breeding scheme.
- In most circumstances, sexing of either semen or embryos appears to be of little value in accelerating genetic improvement. However, the development of a cheap, reliable technique for sexing semen in large enough quantities for conventional AI, has led to improvements in the dissemination of genetic improvement and in the efficiency of animal production. Embryo sexing on a smaller scale could still allow more effective dissemination if it is coupled with *in vitro* production of embryos.
- Cloning appears to be of fairly limited value in accelerating genetic improvement, but the potential of the technique to accelerate dissemination of genetic improvement to commercial herds or flocks is clear if the techniques become more cost effective and efficient. Surrogate sire technology, where elite donor germplasm is disseminated via germ-cell ablated surrogate males, also has major potential to achieve similar impacts.
- The rate of genetic progress is influenced by how early in life phenotypes for traits of interest can be recorded, and by the quality and quantity of data available to predict breeding values. Several technologies now exist to allow (automatic or other) recording of

hitherto difficult-to-measure performance traits (e.g. traits associated with disease, reproduction, feed intake, carcass composition and greenhouse gas emissions) or enable recording of such traits earlier in life. These include the use of milk mid-infrared spectral data to measure various performance traits in dairy cattle, sensors for measuring disease and fertility-related traits, automated feed intake recording systems, video imaging and a range of scanning techniques for predicting live animal or carcass composition, and several methods for predicting methane production.

3.2 Study book



Title: ***Pig Production***
Authors: *Andreas Vest Weber, and
Maria Eskildsen*
Year: *2016*
Publisher: *SEGES*

3.2.1

Weber and Eskildsen (2016)

Page 175-195

9. Pig Breeding



Denmark is world champion in pig breeding. The Danish breeding system DanAvl is a unique and very efficient system which ensures that pig producers always breed on the best genes. The Danish breeding system, along with good production management, helps ensure that Danish pig production holds on to the leading position in the world.

DanAvl has existed since the end of the 19th century and is managed by SEGES Pig Research Centre. DanAvl has a board that consists partly of representatives of pig producer associations in Denmark and of the associations "Danske Slagterier" (Danish Slaughterhouses), "Dansk Landbrug" (Danish Farmers Association), "Danske Svineproducenter" (Danish Pig Producers), and partly of pig producers, who have joined of their own volition.

All parties involved in the DanAvl system are obliged to adhere to common rules that are established by SEGES Pig Research Center. Outside of the pure breeding, the people involved in the DanAvl system also perform activities related to marketing, sales and distribution of breeding material.

Parties involved in the DanAvl system:

- *Genetic management*
 - SEGES Pig Research Center
- *Artificial insemination companies*
- *Distributors*
- *Multiplier herds*
- *Nucleus herds.*

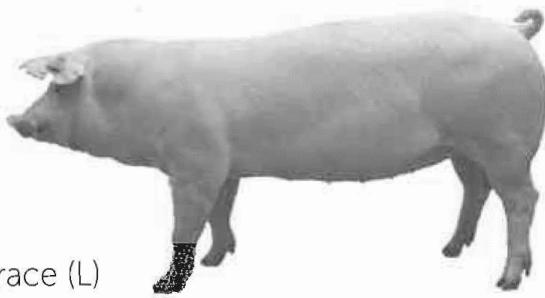
Breeds

DanAvl uses three breeds in the breeding program: Landrace (L), Yorkshire (Y) and Duroc (D). When abbreviating cross-breeds, the male breed is always mentioned first. The designation LY thus indicates an offspring from a Landrace boar and a Yorkshire sow, and L(YL) is the offspring from a Landrace boar and a YL sow.

Due to high fertility, litter performance and excellent maternal qualities of the Landrace, the sows are used as dams of YL young females which together with LY young females are the most suitable hybrid sows for producing finishing pigs.

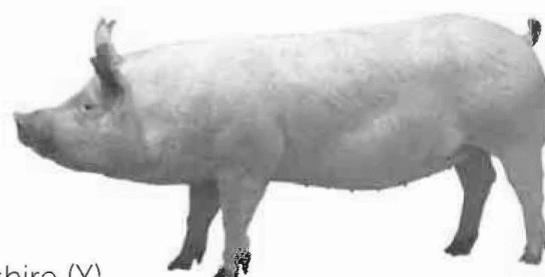
Yorkshire, like Landrace, is used as dam of LY young females. It is also the most suitable breed in relation to production traits. Danish Yorkshire has a high meat percentage, high daily gain, low feed conversion ratio and a good meat quality. In addition, Yorkshire has good litter performance and maternal qualities.

Duroc originates from the United States and Canada and was imported in 1977-1979. Subsequently, it was used as a sire line for cross-breeding. Since then, Danish Duroc has been improved, especially regarding meat percentage and daily gain. Duroc produces fast-growing finishers with a low feed conversion ratio and a high meat percentage.



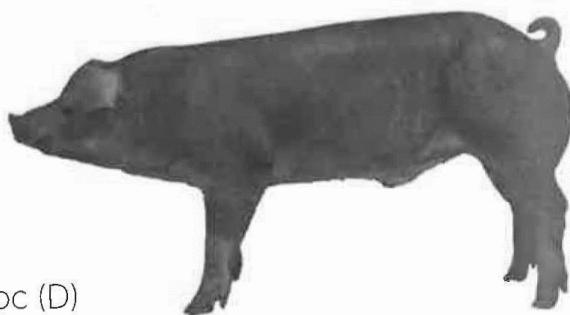
Landrace (L)

White breed with long, narrow and light head with large drooping ears. The body is long with a light front and full hindquarters. Landrace is used as a dam line. Very high litter performance and excellent maternal qualities make Landrace sows very suitable for dams of YL hybrid sows. Crossed with Yorkshire or zig-zag crossbreed sows, Landrace boars are used for production of hybrid sows.



Yorkshire (Y)

White breed with short and broad head and protruding ears. The body is medium long, low and the shoulder region is a little heavy. Yorkshire is used as dam line. This makes the sows of this breed suitable as dams of LY hybrid sows. Crossed with Landrace and zig-zag crossbreed sows, Yorkshire boars are used for production of hybrid sows.



Duroc (D)

The breed is chestnut brown, and the colour varies from totally blond to almost black. It has a short, broad and heavy head with small dropping ears. The body is short and broad with a little coarse front and full hindquarters. Duroc is used as a sire line and has a particularly high daily gain and an excellent feed conversion ratio. Duroc boars are of great economic value in the production of finishing pigs and are thus used for that.

In addition to that, Duroc produces carcasses with good meat and eating quality. Today, DanAvl has the largest population of pure-bred Duroc in Europe.

9.1. Landrace, Duroc and Yorkshire are the three prevailing breeds in Danish pig breeding.

Thus, Landrace and Yorkshire are used as dams, and Duroc is used as sires. By crossing LY/YL females with Duroc semen, pig breeders obtain complete heterosis, excellent meat quality and production results. This crossing produces a D(LY) or D(YL) finishing pig.

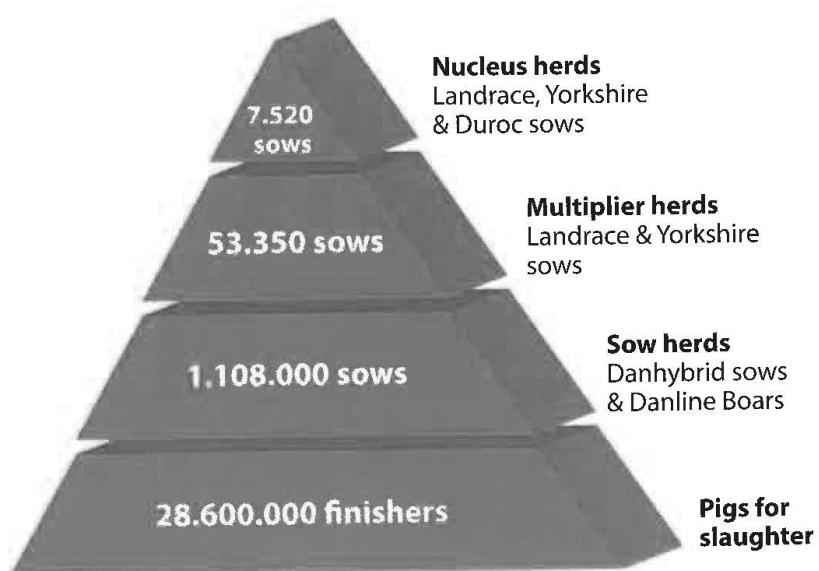
er herds which cross Yorkshire young females with Landrace semen and Landrace young females with Yorkshire semen. The multiplier herds thus produce LY or YL crossbred animals. The females from these crosses are sold to the production herds where they are inseminated with purebred Duroc semen from the AI-stations.

Breeding Structure in Denmark

The breeding structure roughly consists of a small group of breeders who produce purebred Yorkshire and Landrace young females and Duroc boars. The Duroc boars are transferred to AI-stations where they produce semen primarily for production herds.

The purebred Yorkshire and Landrace young females are produced in multipli-

SEGES Pig Research Center has agreements with 26 Danish breeders who together have 38 herds with purebred breeding stock: 12 Duroc herds, 13 Landrace herds and 13 Yorkshire herds. Some breeders, thus, have several breeds. As of August 2015, 134 multiplier herds are approved, of which 28 are associated with a breeding herd. In addition to the Danish breeding and multiplier herds, SEGES Pig Research Center has engaged with an increasing number of foreign herds.



9.2. Breeding structure in the DanAvl system. Moreover there were 83 foreign multiplier herds (August 2015).

Testing of Breeding Stock

On DanAvl's homepage, www.DanAvl.com, it is possible to find the family tree and indices of all registered animals since 1985, if only the ID-number of the animal is known. Moreover, the page contains a ranking of breeders as well as multiplier herds, and there is a hit list of available boars currently housed at AI-stations. On the homepage it is possible to order an electronic breeding and multiplier report which provides access to information on inbreeding, among other things. To learn about inherited traits, a thorough registration and testing of every single breeding animal is carried out.

Testing is carried out partly in the breeding herds, partly at the test station Boegildgaard where selected boars are inserted at weaning.

Testing is carried out while the animals weigh between 30 and 110 kg. The animals are fed ad libitum during their entire growth. There are requirements to the used feed compounds and dispensers. In addition, the animals must be housed gender-divided.

Every week the breeding herds and Boegildgaard are visited by a breeding assistant from SEGES Pig Research Center. The animals are weighed, back fat scanned and their conformation is assessed. Together with the data from relatives, these data form the basis of animal subindices of daily gain, feed conversion ratio, meat percentage and conformation.

All other data from the breeding and multiplier herds are registered and sent to the data bank of pig breeding. Data comprises inflow and outflow of sows, litter and

9.3. At Boegildgaard it is possible to register the feed intake of the individual animal. This way the individual feed conversion can be calculated. The best boars have a feed conversion ratio of about 1.90 feed unit/kg daily gain.



Source: SEGES Pig Research Center

9.4. The gene flow at DanAvl. About 90,000 animals are tested in breeding herds every year, while further 3,500 boars are tested at the test centre Boegildgaard annually.

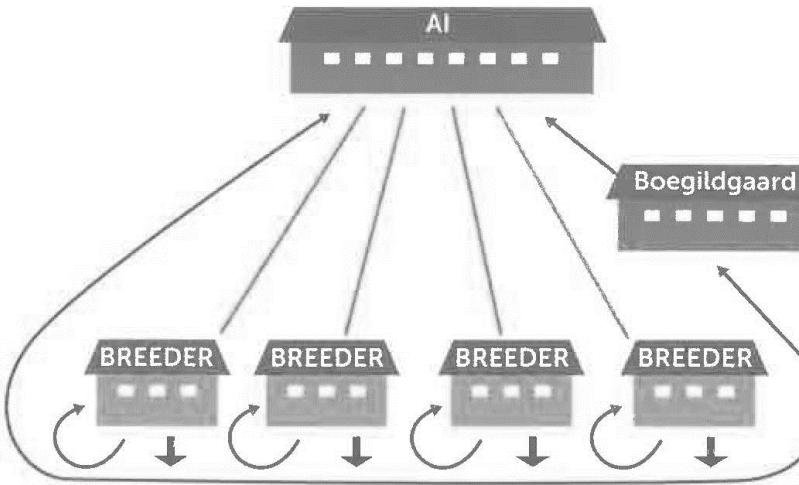


Illustration: Lene Kuse Køsset
after SEGES Pig Research Center

insemination registrations, weighing in of breeding animals and in-herd performance testing.

Every Thursday night, a national index of the entire breeding stock is calculated including registrations from all breeding, multiplier and test station herds.

Parallel to the in-herd performance testing, the test station Boegildgaard conducts an in-station testing of young boars. The animals for in-station testing are selected in the breeding herds at the age of four weeks, among litters with the highest indi-

ces that week. The boars are transferred to a climate unit where they are penned for five weeks. The animals are tested for the same characteristics as the in-herd tested animals. However, all pens are supplied with a transponder station enabling computer registration of the feed intake of the individual animals.

At the end of the testing period, the best boars are selected for the AI-stations. All other boars are slaughtered. In connection with slaughter, meat quality characteristics are registered.



9.5. Duroc piglets suckling a Landrace sow in a Duroc breeding herd. The white breeds have better maternal qualities and are therefore very suitable as nursing sows

Danish Breeding Stock Abroad

In 2014 the following number of pigs was sold:

- 4,700 purebred young females for sale in Denmark
- 22,000 purebred young females for export
- 258,000 cross-bred young females for Danish herds
- 440,000 cross-bred young females for export

In addition to Danish breeding and multiplier herds, SEGES Pig Research Center has an agreement with an increasing number of foreign multiplier herds. Currently, there are 83 foreign multiplier herds with Danish breeding stock.

- 5.3 million doses of semen for sale in Denmark
- Approximately 1.8 million doses of semen for export
- 450 boars sold in Denmark
- 3,350 boars sold abroad.



9.6. DanAvl market share is developing favourably. The map shows all countries supplied with DanAvl breeding material worldwide.

Breeding objectives

The characteristics strived for in breeding are called breeding objectives. As a starting point, the breeding objectives in DanAvl are established out of consideration for the economy of the pig producers and the slaughterhouse. SEGES Pig Research Center establishes the breeding objectives of the three breeds in DanAvl.

Some characteristics have a high heredity and are thus easy to improve through breeding. This applies to daily gain, feed conversion ratio and meat percentage. Fertility and strength have a lower heredity; the improvement of these characteristics on the other hand, has a very high economic value.

Generally, the more characteristics are included in the breeding objectives, the lower the progress per characteristic. The total economic value of progress in the char-

Characteristics in the Breeding Objective

Must comply with the following requirements:

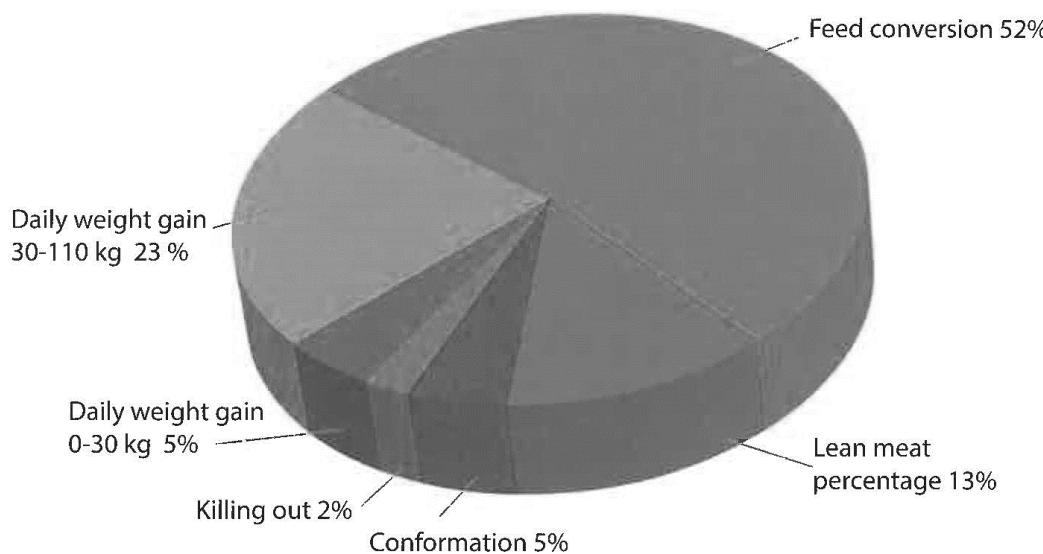
- They must be of production economic importance
- The economic value of the characteristics must be definable
- They must be hereditary
- They must be either directly or indirectly measurable.

acteristics of the breeding objectives may, however, increase. Therefore, the breeding objectives are made up of the combination of characteristics which provide the highest possible economic value for the pig producers.

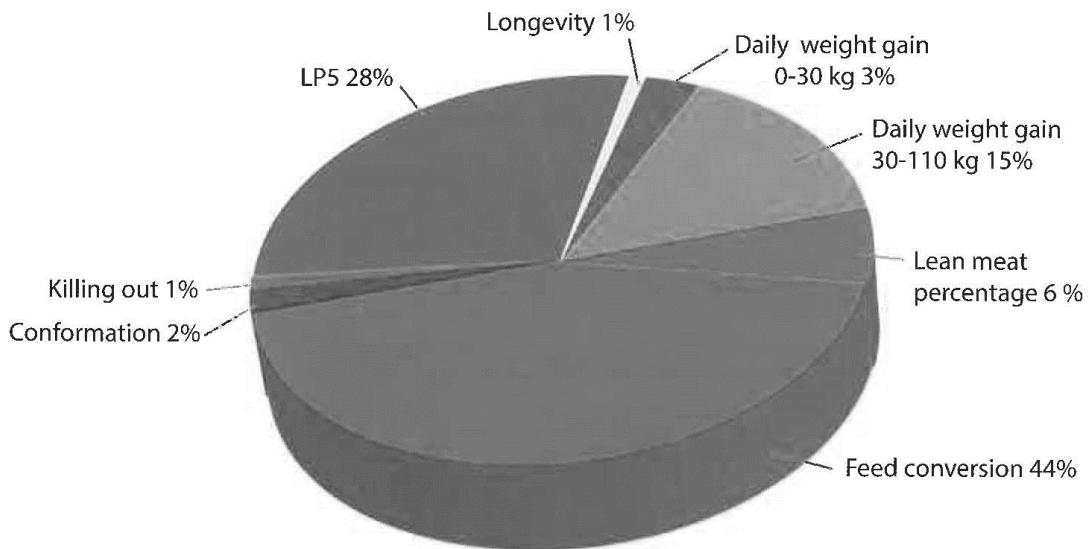
There are different breeding objectives for Duroc and the female breeds Landrace and Yorkshire, since longevity and high fertili-

9.7. Breeding objectives in DanAvl.

Daily weight gain from 7-30 kg	Measured in gram/day.
Daily weight gain from 30 -110 kg	Measured in gram/day.
Feed conversion	Feed units/kg gain.
Lean meat percentage	Meat percentage is measured by ultra sound scanning in tests and in the carcass of slaughtered boars from Boegildgaard.
Conformation	Subjective exterior assessment made by breeding technician. The animals are graded from 0-5.
Killing out	Measured in percentage on carcasses for boars that have been tested.
Longevity	Probability that a sow which has had the first litter will also be inseminated for the second litter.
LP5	Number of live piglets/litter 5 days after farrowing.



9.8. Characteristics of the boar breed Duroc and distribution of genetic progress (measured as economy).



9.9. Characteristics of the sow breeds Landrace and Yorkshire and distribution of genetic progress (measured as economy).

ty is not important in Duroc which is only bred for its characteristics as finishing pigs.

Every third to fourth year the economic values are revised, and it is decided which characteristics should be used for breeding. If, for example, feedstuff is very expen-

sive over some time, it will be economically advantageous to have a good feed conversion, and therefore the economic emphasis on feed conversion is increased, hence the animals with a good feed conversion will have a higher index.

Index

All animals in DanAvl have an index which is an estimate of the genetic level of the animal. The index is composed of subindices of the different characteristics in the breeding objectives multiplied by the economic value of the characteristics.

The subindex is the deviation from the average of all active progeny tested breeding sows in the breed and the sows dead or slaughtered within the last six months. The subindices can be compared with the individual categories of modern pentathlon. The winner is the overall best in all five categories, no matter if others are better in one or two disciplines.

The average index among breeding sows is always 100. The higher above 100 a breeding animal is, the higher is the value of the animal. Weekly, the indices of all breeding stock in the breeding system are recalculated (also those of dead animals), so the latest registrations are currently used. On DanAvl's homepage DanAvl.com, it is possible to find the index of an animal by using the ID-number of the animal.

After the weekly index calculation, SEGES Pig Research Center selects the best boars and they are transferred to the AI-stations. The boars originate from either the test station Boegildgaard or directly from the breeding herds.

The breeders select the females in the herd with the highest index and use the best boars available in the entire system, in order to produce the next purebred generation of breeding stock.

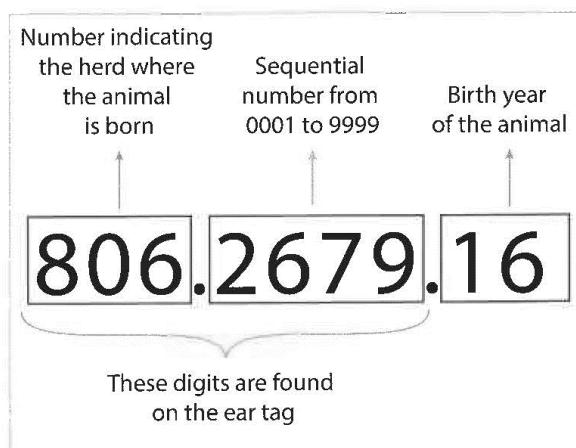
Semen is distributed so that breeders have equal access to semen from top boars within the three breeds. To reduce the risk of inbreeding, there is a quota for how many litters a boar is allowed to sire in every breeding herd.

ID Number

All breeding stock in DanAvl has a nine digit ID number. The first seven digits are indicated on the breeding ear tag.

Before weaning the breeding animal is ear tagged with its individual ID number. The three-digit herd number is also indicated on the ear tag. The ID number can be created in various management systems for use in nucleus management. Ear tags are ordered through DanAvl and the colour indicates the breed in nucleus management:

- Landrace: blue
- Yorkshire: orange
- Duroc: white.



9.10. ID number.

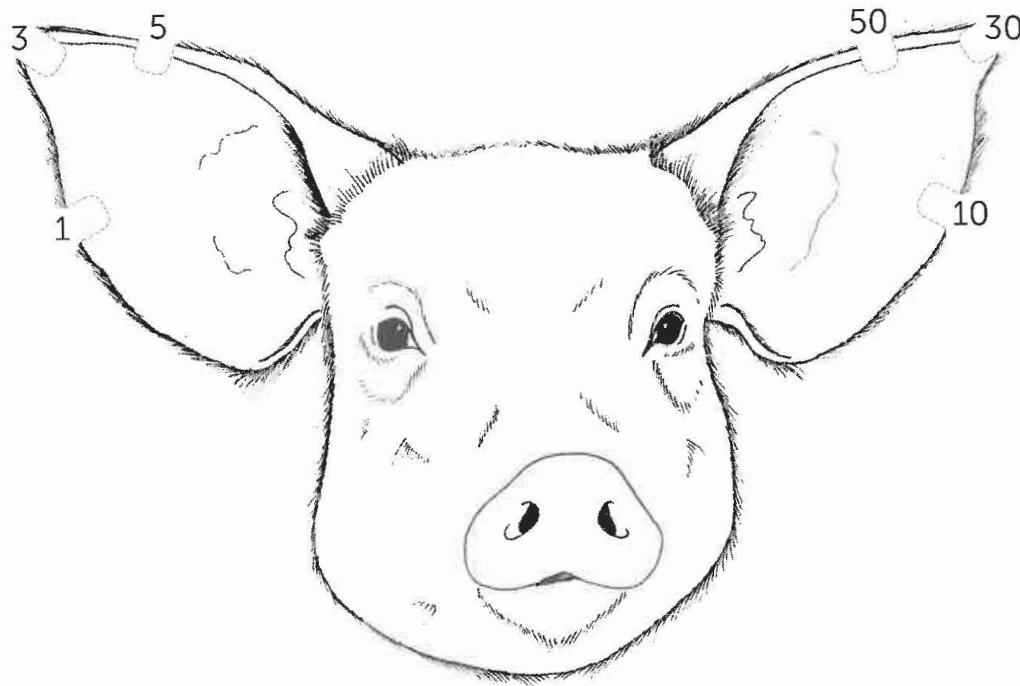


Illustration: Chanett Israell Colmorn

9.11. System for ear notching of young animals. These notches always allow checking the age and descent of the pig.

At birth, the selected young females may be notched with the last two digits in the sow-number of the dam.

Genomic Selection

DanAvl uses genomic selection in all breeds and of all characteristics. At genomic selection, the DNA of the animals is tested enabling prediction of aptitude for daily gain, meat percentage etc. at an early stage in the life of the animal. This way, the underperforming animals can be culled earlier and breeding progress can be accelerated. A breeding system using genomic selection is, in principle, not very different from a normal breeding system. The breeding candidates must still be tested, and a breeding index must be calculated. This breeding index is used for selecting which animals to DNA-test, as it is not economically feasible to DNA-test all



Photo: Jens Tønnesen, LandbruksMedierne

9.12. At weaning, at the latest, the young female gets its individual ID tag. Breeding and multiplier herds must use yellow ear tags.



9.13. Removal of hair sample for DNA analysis.

animals. The DNA-test subsequently enables calculation of a more exact index, indicating whether to use the animal for breeding.

Genomic selection:

1. Testing of potential breeding candidates
2. Calculation of breeding index
3. DNA-test based on breeding index
4. Calculation of breeding index using DNA information
5. Deciding whether to use the animal for breeding.

9.14. AI stations have different boars, producing semen for different purposes.

Production semen	
Duroc	Ordinary production semen where semen from up to ten different Duroc boars is mixed and used for production of finishing pigs.
Duroc Plus	Boars with an extra high index for meat percentage. Duroc Plus is production semen from selected boars with a subindex for meat percentage of at least 0.6 % above average. Duroc Plus is expected to give the finishing pigs a meat percentage which is 0.3 % higher than ordinary Duroc production semen.
Top Duroc	Selected with a higher average index. On average, the indices will be about one point higher than ordinary Duroc boars.
YD boars	The semen is used for production of finishing pigs. The influence from Yorkshire produces offspring with a good meat percentage and only a few with coloured markings. The animals are white due to dominance of Yorkshire.
Pietrain	Imported boars from Germany. Pietrain is used in many European countries and is also known for producing large hams and a high meat percentage.
Name semen	
Landrace	From a specific, named Landrace boar. The semen is used for production of hybrid sows on Yorkshire and for zig-zag cross-bred sows.
Yorkshire	From a specific, named Yorkshire boar. The semen is used for production of hybrid sows on Landrace or zig-zag cross-bred sows.
Top name semen	From named boars with high indices. Available from Landrace, Yorkshire and Duroc. The semen is primarily used for production of purebred animals.

On-farm AI

On-farm AI is artificial insemination with semen from boars in own herd. Before a new boar is used for AI or mating in a herd, the quality of the semen should be checked. It is done by inseminating ten animals, neither young females, weaned nurse sows nor returners, with semen from the boar. If at least seven of the sows subsequently are gestant, the boar is "conditionally approved" for on-farm AI. If less than seven sows become gestant the boar cannot be approved yet, and further ten sows are inseminated. If again less than seven sows become gestant, the boar is slaughtered.

Equipment for On-farm AI

As a mounting base for the boar a phantom is used. The phantom must be covered with soft washable material without tears in the coating. The mounting base must be nonslip and fastened to the phantom to prevent it from moving.

The semen is collected in an insulated container, lined inside with a plastic bag which is everted over the container edge. The lined bag and gauze is held in place by a strong rubber band. Before the bag is used it must be filled with air to prove close.

Containers in direct contact with semen must not be made of metal or glass, as they are too cold for the semen. If semen is exposed to cold shock, fertility will be impaired. All material used more than once must be of plastic. All other material should be scalded with boiling water and

Collection of Semen:

- Phantom, gloves and container are taken into the pen.
- Phantom is fastened to the pen wall in order to be stable when the boar mounts.
- Both gloves are put on when emptying prepuce.
- When prepuce is emptied, one glove is taken off and the inner glove is ready for the actual collection.
- By massaging the prepuce forward, the boar is stimulated to ejaculate. The boar is also stimulated by the sight and smell of the phantom.
- The tip of the penis is grasped firmly so that the uttermost tip rests on the edge of the hand. When the hand is clenched around penis it is able to screw stuck, like it does in the womb of the sow.
- The first 10-15 ml should not be emptied into the container, as they contain urine residues and pus and very little sperm per ml.
- Penis is held upwards to prevent urine, secretions and pus from draining into the semen. Penis must be held in a gentle curve upwards to allow the semen to drain into the container. Penis must not be bent.
- In the last phase of the collection a "gel-plug" appears. It lies on top of the gauze and should not be collected.

subsequently dried out completely. Water kills sperm. If equipment has been washed in a dishwasher it may contain soap residues which will also have a spermicidal effect on the semen.

In order to ensure a hygienic collection, the container must be covered by two disposable gloves. They must be made of talc-free vinyl as latex and talc have a spermicidal effect.

Transfer of boars requires special attention because they are heavy and may be aggressive if they feel under pressure. It is crucial to train boars several times a week before they are grown. Boars must be handled consequently, predictably and quietly.

The boar should be given sufficient time before collection is stopped. If collection is stopped too early, the libido of the boar may be reduced. The boar dismounts the phantom without help when it is finished, otherwise the fact that the "gel-plug" has come out is a sign that it has finished. Collection may last five to ten minutes from ejaculation until the last semen and the gel-plug has come out.

Assessment of Semen

Boar semen has a neutral smell. If it smells strongly, it is contaminated and must be discarded. Semen must not be discoloured. Blood in the semen gives a brownish colour. Blood damages sperm, causing impaired fertility. If semen separate, it is contaminated with pus or bacteria. Semen must be discarded if at all in doubt about the quality. Urine in the semen causes it to be yellowish.

A dose of semen may vary from 80-600 ml. There must be a sufficient number of viable sperm in every ready-mixed dose, and therefore semen concentration is assessed. It is, however, not possible to do it exactly without a counting chamber.



Photo: SEGEC Pig Research Center

9.15. It is crucial to focus on hygiene and be thorough when diluting and assessing the motility of the semen in the microscope.

Counting chambers are usually only profitable at AI stations.

Instead the semen may be assessed in the colour:

- Blue-grey skimmed milky = only few sperm cells
- Milky white = medium number of sperm cells
- Cream-coloured = many sperm cells.

Motility is an expression of sperm forward movement (mobility) and normal appearance (morphology). Sperm cells swimming backwards or in circles are not assessed to be fertile. If there are drops of cytoplasm on the sperm cells, semen is not fertile. Cytoplasm is a sign of immature sperm cells and can be caused by too frequent collection of semen from the boar.

As soon as possible after collection and filtration, semen must be diluted by a special boar semen diluent containing antibiotics. Semen must be diluted minimum 1½ times an hour after collection at the latest. If semen is left undiluted for a long time, it will be damaged. The diluent contains nutrients which are necessary for the survival of the semen. It is important that the diluent is thoroughly warm. Temperature must be between 30 and 35° C.

At AI stations diluting is carried out step by step. In the first step, a pre-layer is made by diluting the fresh semen by 30° C warm diluent an hour after collection at the latest. On average, the pre-layer must contain two billion motile sperm cells per 25 ml. In the second step, the pre-layer is diluted by diluent until a final volume of 80-85 ml is obtained. Temperature in the final diluent must be between 20 and 30° C.

Diluent which has been opened must be used within 72 hours. Diluent must be kept in a refrigerator until needed for use.

It might be an advantage to buy diluent doses of several sizes for use on busy and not so busy days of insemination.

A drop of the diluted semen is placed on a slide on the hot plate (37-39° C) in the microscope. Place a cover glass on top. The morphology of the sperm cells must be assessed. If there are more than 30 percent defect sperm cells, the semen is discarded.

It takes roughly half an hour to collect, assess and dilute a dose of semen.

Replacement of Sows

Annual culling of sows is about 50 percent, meaning that up to half a million sows are replaced in Danish herds every year. It is vital to select the right young females for replacing the culled sows.

9.16. Only semen with more than 70% fertile sperm cells can be used.



Photo: Hatting KS

9.17. Reasons for culling sows in Danish herds.

Reasons for slaughter	%
Not gestant	28
Poor maternal qualities	25
Other – age	20
Disease and injuries	14
Lack of oestrus	7
Too many in the weekly batch	5
Behaviour	1
Total	100

Source: SEGES Pig Research Center

Home-bred Young Females

Herds have two possibilities of recruiting young females. Either produce them in the herd or purchase them from multiplier herds. By home-breeding young females, the herd does not risk external infection and thus it is not necessary to have quarantine facilities. In turn, it takes time and passion for breeding and additional space for the breeding stock. In addition, home-bred young females leave a residue of males which only have genes from white breeds and not Duroc. They will have poorer finisher characteristics than males cross-bred with Duroc.

Young females are produced in a purebred nucleus herd or by zig-zag-crossing.

The herd has the purebred nucleus as its own little breeding herd in the herd. The nucleus usually constitutes 10-15 percent of the year sows and consists of purebred Yorkshire or Landrace sows which are inseminated with name semen or top name semen in order to produce YL or LY young females. The nucleus must be maintained by selecting the best nucleus sows and inseminate them with purebred semen.

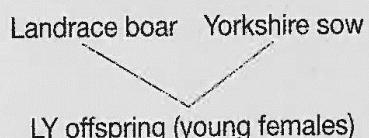
By zig-zag breeding, all sows in the herd are potential breeding stock, and the animals with the highest index are selected for breeding. By zig-zag crossing, semen from respectively Landrace and Yorkshire is used.

Cross-breeding Methods

Different cross-breeding methods are used to produce young females and finishing pigs, as the case may be.

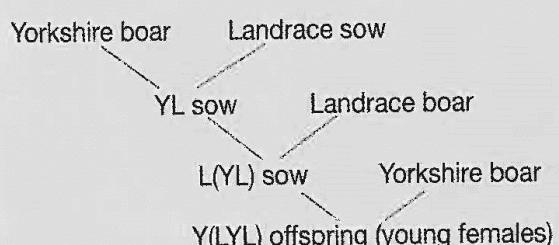
One-way crossing

Offspring heterosis is optimal, i. e. increased litter size, more obvious signs of oestrus and gestation is more easily achieved.



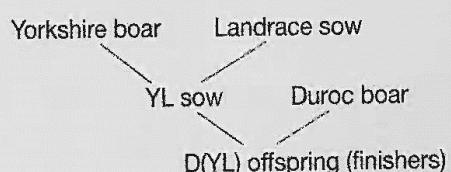
Zig Zag crossing

Boars from two different breeds are used alternately.



Three-way crossing

Heterosis of the sows is maximized, and by using boars of a third breed heterosis is also maximized in the offspring. Is used for production of finishing pigs.





9.18. Ear tagging of female pigs.

Selection of Young Females

It is crucial that young females are treated as breeding stock and not as finishing pigs. No matter which breeding method is used, it is vital for production to select the right animals for breeding, and to cull and sell the remaining pigs as finishers.

At birth, female pigs weighing at least one kg and with minimum 14 teats are selected. Young females with a high birth weight are expected to have larger litters.

To reduce wear damage on the tip of the teats, an adhesive plaster is put on the front teats down to the navel. The navel must not be covered. Usually, the plaster falls off, otherwise it is removed on day two or three.

To ensure that the future breeding stock gets the necessary nutrients, litters with future breeding stock must be litter equalized to maximum 12 piglets.

Young females must also be discriminated positively after weaning. Due to difference in temper and behaviour, it is recommended that young females are penned together and not mixed with finishing pigs.

It is very important that young females get used to human contact. Hence, staff must go into their pens minimum twice a day. In addition, it is an advantage to pen young females in a part of the herd with much activity.

Culling of Female Pigs

At farrowing, seven female piglets are expected per litter, on average, but only few of these animals are suitable for becoming

ing production sows. Up to seven and a half kg, there will be a dropout of 30-40 percent, leaving 4.2-4.9 potential breeding animals per litter at weaning.

Minimum four times between weaning and 100 kg, a critical culling of young females unsuitable for becoming sows must be carried out. Reasons for culling are unthriftiness, wrong leg position, hernia, faulty hooves and saddle, to men-

tion some. The first culling can be done at transfer to the weaner unit where a culling percentage of two and a half to six is expected.

Again at transfer to the finishing unit up to four percent should be culled, leaving 3.8-4.7 suitable young females per litter at 60 kg. Culling as early as possible provides additional space in the pen for the remaining young females.



Photo: Tens Tens - Agri Lantbruk Mæleme

9.19. Young females must be penned where there is much light and frequent contact with staff. This way they get used to human contact and will consequently be easier to handle for vaccination and oestrus control.



9.20. Only the very best female pigs will be used as future sows.

Young females are thus assessed at every transfer to a new unit, and they must be assessed at least two times more after 60 kg. The first time should be before the culled young females exceed the weight limit for finishing pigs. The second assessment should be done when recording the first oestrus; here there should be at least 13 functioning teats. Special attention should be paid to the position of the teats, inverted teats or short and thick teats with a recess at the tip of the teat.

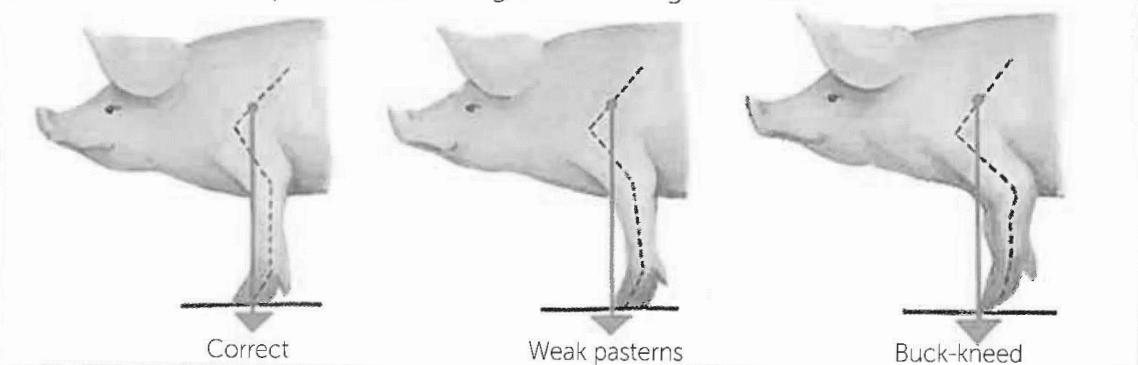
Young females are culled if they have failed to show oestrus at the age of 11 months. If the young female has not reached oestrus within 11 months, it is not likely to become an efficient sow. At 90-110 kg, at the latest, young females with critical defects on legs and hooves are culled.

The number of breeding litters should be large enough to eliminate half of the female pigs before insemination. Alternatively, it may be necessary to inseminate animals with insufficient longevity or older sows that should have been culled. Both will have an adverse effect on productivity.

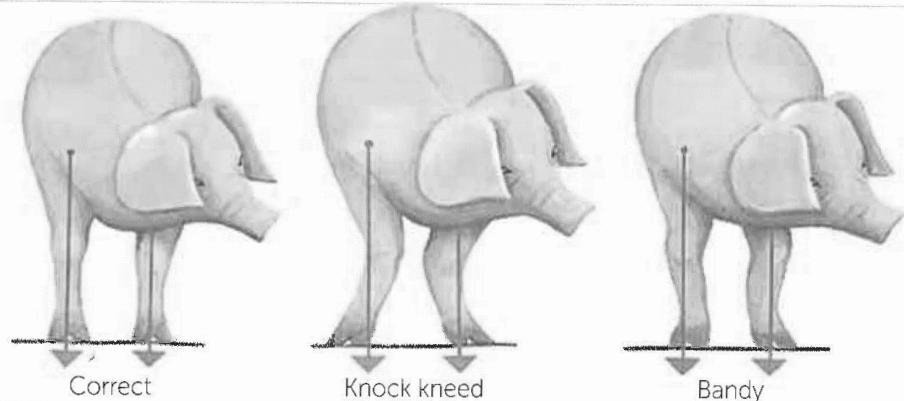
Nucleus Management

When home-breeding young females, it is important to keep a high index level on the animals, otherwise the herd risks being side-tracked when it comes to breeding. For this, the programme Nucleus Management may be used.

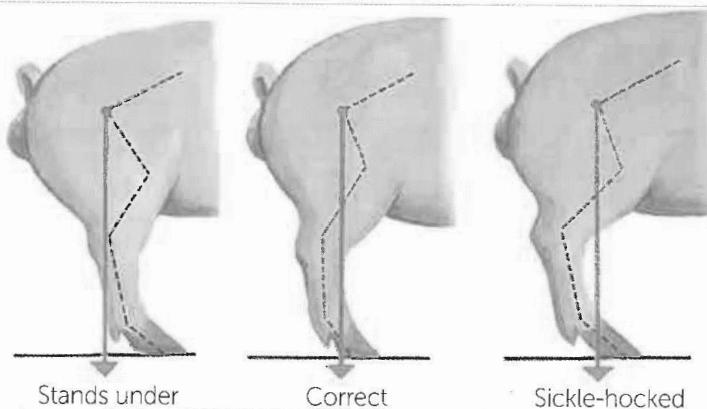
9.21. Assessment of pasterns (front legs) in breeding animals.



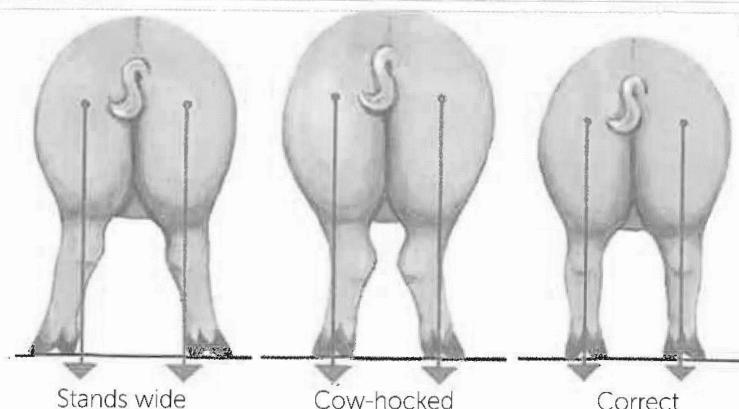
9.22. Breeding assessment of pasterns.



9.23. Breeding assessment of hocks (hind legs).



9.24. Breeding assessment of hocks.



9.25. Advantages of two different recruitment strategies when recruiting young females.

Home-bred	Purchased
The herd is kept closed	Pens for young sows not required
Quarantine facilities not necessary	Space and nucleus management not necessary
No agreement on regular supply of breeding stock	No purchase of name semen
	Faulty animals are replaced

The nucleus management programme continuously updates the index of the sows in the nucleus or zig-zag herds, when the sows in breeding and multiplier herds have their indices updated. The use of nucleus management requires reporting of influx, inseminations, farrowing and exit of all breeding stock and ear tagging of all home-bred animals.

9.26. If there is no cohesion between age and weight of the young female, it may indicate that the animal has been ill or failed to have sufficient nutrients during growth. Such animals are not suited for being future breeding stock. Hence, it is important to know the age of the young females so they can be transferred to the insemination unit according to age and not size.

Purchased Young Females

If young females are purchased from a breeding or multiplier herd, it is important to assess the animals thoroughly on arrival in the herd. The position of the legs, the hooves and the general appearance of the animal should be checked. The young female should not have open wounds after tail biting, ear or flank suckling or the like. Wounds must be reported the day after receipt at the latest. Complaints about too short tails must be submitted within eight days after receipt, and reports on legs, hooves and general exterior characteristics must be submitted within 20 days after receipt. Last but not least, it is, of course, of vital importance for production that the purchased young females have a high index.

Age, days	Weight, kg						
8	6.5	75	25	127	53	183	93
49	14	85	30	141	63	197	99
63	20	99	37	155	73	210	105
		13	44	169	83		

Housing Conditions of Young Females

The choice of penning system for the young females depends on the feeding principle in the herd, and when and how often during growth the young females are transferred.

Young females must always be penned on a nonslip floor and have plenty of space. As the animals grow, space requirements increase. To ensure a normal wear of the hooves, it is crucial to design the pens with partially solid and partially slatted floor. Deep litter is not recommended as it causes too little wear of the hooves. To ensure good reproductive qualities, there should be sufficient lighting. Minimum 100 lux is recommended in the young female unit. 100 lux is equivalent to enabling a person with a normal sight to read a newspaper in sow-height.



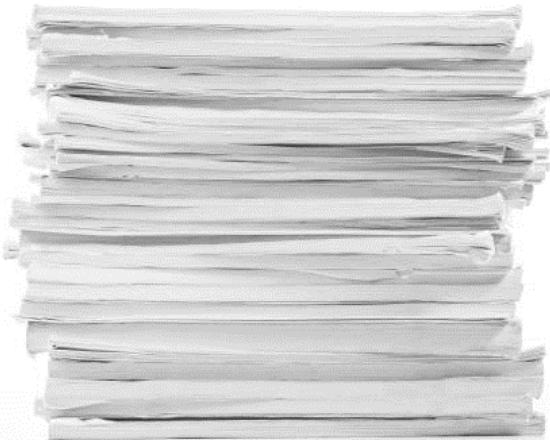
Photo: Jens Tønnesen, LandbruksMiljø

9.27. The number of young females should be large enough to cull half of them between weaning and insemination. Young females over 300 days old and young females of a poor constitution and deviant size compared to age should not be inseminated.

9.28. Young females are future sows in the herd, hence it is crucial to handle them correctly. Young females with strong legs, in good shape and appropriate body condition at insemination will have a longer life as production animals and provide better results. Home-bred young females can be housed in two-climate pens until they weigh about 30 kg, just like the other weaners in the herd. Subsequently, they should be transferred to a separate young female unit where they usually stay until first oestrus is noted.

Weight	Regulatory space requirements per young female	Recommendation per young female
50-85 kg	0.55 m ²	0.75-1 m ²
85-110 kg	0.65 m ²	0.75-1 m ²
110---	1 m ²	1.9 m ²

3.3 Scientific articles



3.3.1

Nielsen, H. M; Olsen, I.; Navrud, S.; Kolstad, K.; Amer, P. (2011). *How to consider the value of far animals in breeding goals. A review of current status and future challenges.* J. Agric. Environ. Ethics 24:309-330.

How to Consider the Value of Farm Animals in Breeding Goals. A Review of Current Status and Future Challenges

H. M. Nielsen · I. Olesen · S. Navrud · K. Kolstad ·
P. Amer

Accepted: 10 May 2010 / Published online: 6 June 2010
© Springer Science+Business Media B.V. 2010

Abstract The objective of this paper is to outline challenges associated with the inclusion of welfare issues in breeding goals for farm animals and to review the currently available methodologies and discuss their potential advantages and limitations to address these challenges. The methodology for weighing production traits with respect to cost efficiency and market prices are well developed and implemented in animal breeding goals. However, these methods are inadequate in terms of assessing proper values of traits with social and ethical values such as animal welfare, because such values are unlikely to be readily available from the product prices and costs in the market. Defining breeding goals that take animal welfare and ethical concerns into account, therefore, requires new approaches. In this paper we suggest a framework and an approach for defining breeding goals, including animal welfare. The definition of breeding goals including values related to animal welfare requires a multidisciplinary approach with a combination of different methods such as profit equations, stated preference techniques, and selection index theory. In addition, a participatory approach involving different stakeholders such as breeding organizations, food authorities, farmers, and animal welfare organizations should be applied. We conclude that even though these methods provide the necessary tools for considering welfare issues in the breeding goal, the practical application of these methods is yet to be achieved.

H. M. Nielsen (✉) · I. Olesen · K. Kolstad
Nofima Marin, P.O. Box 5010, 1432 Ås, Norway
e-mail: hanne.nielsen@nofima.no

S. Navrud
Department of Economics and Resource Management, Norwegian University of Life Sciences,
P.O. Box 5003, 1432 Ås, Norway

P. Amer
AbacusBio Limited, P.O. Box 5585, Dunedin 9058, New Zealand

Keywords Animal welfare · Breeding goal · Economic value · Market experiment · Non-market value · Public good · Willingness to pay

Introduction

Today, European people consider farm animals as an important means for obtaining nutrition, yet key social goals such as food safety, food quality, environmental protection, and animal welfare are also taken into account. Therefore, public authorities are obliged to bear in mind these consumer preferences when formulating and implementing relevant policies regarding farm animal production. There is a growing body of EU legislation on animal welfare that takes into account public concerns, stakeholder input, and possible socio-economic implications (Horgan and Gavinelli 2006). Potential economic benefits of improving animal welfare depend on consumer and citizen belief and interest in an “ethical” production. The consumer perspective reflects people’s own interest in the market, while the citizen perspective reflects people’s opinions, values, beliefs, and concerns with public interests (Sagoff 1988). In general, citizens want a change of agriculture practices while most consumers only incidentally buy environmental and animal friendly products (Kjærnes et al. 2005). To properly assess the potential outcome of costs and benefits of measures to improve animal welfare, it is, therefore, important to understand ethical values and their potential role in both consumer and citizen preferences and behavior.

In regards to selective breeding of farm animals, animal welfare, together with resource use and the effect on the environment, is one of the major issues (Gamborg and Sandøe 2005). Good animal welfare is, among other things, characterized by freedom from pain to disease and the opportunity to express normal behavior (Fraser 2003). Breeding affects animal behavior and animal welfare through genetic changes in traits related to these issues, such as resistance to diseases.

The increasing concern about animal welfare related to animal production means that animal welfare needs to be taken into account in the definition of animal breeding goals, as stressed by a number of authors during the last 15 years (Torp Donner and Juga 1997; Gibson and Wilton 1998; Olesen et al. 2000; Kanis et al. 2004, 2005; Gamborg and Sandøe 2005). With classical index selection, the breeding goal (also called the aggregate genotype) determines the genetic changes in animal traits, and consists of traits weighted by their respective market values (MV):

$$H = MV_1 G_1 + MV_2 G_2 + \dots + MV_n G_n, \quad (1)$$

where H is the aggregate genotype, G_i is the additive genetic value of trait i , and MV_i is the market value (also known as economic value) of trait i , defined by the change in profit of a unit change in the trait (Hazel 1943). Thus, the breeding goal for a given breed or population of animals expresses the targeted direction of genetic improvement for the different traits that are included in the breeding goal and/or selection index.

In the past, livestock breeding programs primarily focused on improving production traits. Selection for production traits only will lead to deterioration of

traits negatively correlated with the production traits; and in some livestock species, it is documented that breeding has lead to increased behavioral problems and health and fertility disorders, thus resulting in lowered welfare of animals (Rauw et al. 1998). An example of this is single trait selection for milk yield, that will reduce resistance to mastitis (see review by Heringstad et al. 2000), and other diseases in dairy cows (e.g., van Dorp et al. 1998), if no proper measures are taken in the selection program to prevent it. Lowered resistance to disease will increase the risk of becoming infected, which again negatively affects animal welfare. Thus, negative side-effects of selection for production traits are in conflict with animal welfare (Sandøe et al. 1999). However, by including traits related to animal welfare in the breeding goal and selecting animals using a multiple trait index, as has been implemented for decades in Scandinavian countries (Philipsson and Lindhé 2003), reduction in traits related to animal welfare can be avoided. Nevertheless, including traits related to animal welfare in the breeding goal may not be enough to avoid deterioration of these traits because many of these traits are unfavorably genetically correlated with production traits (Christensen 1998; Nielsen et al. 2006).

The methodology for weighing production traits such as growth and milk yield with respect to cost efficiency and market prices are well developed and implemented in animal breeding goals. MV for these traits can be derived using profit equations (Brascamp et al. 1985; Groen et al. 1997; Dekkers and Gibson 1998), assuming the goal is the maximization of farmer profit based on the market economy. However, these methods are inadequate for assessing proper values of traits with social and ethical values. This is because values of public goods such as improvements in animal traits, e.g., related to animal welfare, may not be readily available from the product prices and costs in the market, and thus are difficult to value in monetary terms. As with MV, values for improvements in animal welfare will differ in different markets. In addition, assigning values to traits related to animal welfare continues to be a challenge, since the value of such issues may not be transferred via the market (McInerney 2004).

Defining breeding goals for sustainable production systems taking animal welfare and ethical concerns into account will therefore require new approaches. The challenge in defining the breeding goal is to value such intangibles as animal welfare in monetary terms. Olesen et al. (2000) suggested including a “non-market” value in addition to the MV in the breeding goal to consider values related to animal welfare and social aspects such as environmental concerns. Star et al. (2008) introduced the concept of robustness in poultry, which includes individual traits relevant for health and welfare. They suggested that applying robustness in the breeding goal will increase the animal’s ability to interact successfully with the environment.

Tools for deriving values related to animal welfare and other social aspects have been proposed (Olesen et al. 1999; Lawrence et al. 2004). This includes methods based on individual and decision makers’ preferences as well as methods based on selection index theory (Hazel 1943). However, literature on how to derive values to traits related to animal welfare (e.g., Kanis et al. 2005; Nielsen et al. 2005, 2006) is scarce.

Therefore, the objective of this paper is to outline challenges associated with including welfare issues in breeding goals in modern farm animals and to review the

current status of available methodologies and discuss their potential advantages and limitations to address these challenges. Finally, we suggest a framework and approach for defining breeding goals taking animal welfare properly into account.

Characteristics and Valuation of Animal Welfare

Characteristics of Animal Welfare

As pointed out by several authors (e.g., Sandøe et al. 2003; McInerney 2004), there exist several different definitions of animal welfare. Broom (1991) for example, defined animal welfare as the animal's state in regards its ability to cope with its environment. Fraser (2003) identified three different views on how animals should be raised to ensure good welfare and how welfare should be judged. Firstly, animals should be raised under conditions promoting good biological functioning with respect to growth, health, and reproduction. Secondly, animals should be raised in a way to minimize suffering and promote contentment; and finally, animals should be allowed to lead relatively natural lives. Sandøe et al. (2003) stressed that the scientific assessment of animal welfare should be followed by ethical questions about standards for morally acceptable animal welfare, such as what a good animal life is and which compromises are acceptable. The definition of animal welfare will not be further discussed here. In the paper we use different traits related to farm animal welfare as resistance to diseases as examples (e.g., resistance to mastitis in dairy cows). However, we do not deal further with aspects associated with the traits such as how disease resistance should be measured and we do not consider specific traits and aspect such as if the disease is metabolic or infectious.

In the following paragraph, we put animal welfare into an economic context in order to set the basis for how to value animal welfare. Economists define goods according to whether they are private or public goods. As will be illustrated below, in principle, animal welfare can be both a public and a private good but is mostly considered to be a public good. Private goods are characterized as being both exclusive and rival (Randall 1987). Where exclusive goods are those by which payment promotes access to their use or consumption, with a lack of payment preventing use or consumption; and rival goods being those by which simultaneous use or consumption by multiple consumers cannot occur (e.g., bread and meat). Private goods affected by an improvement in, e.g., animal growth, include for instance, quantity of meat and nutrients, for which the value can be traced in the market price. The value of improvements implying lower costs per unit of product, e.g., feed costs due to improved feed efficiency, can also be derived from market prices of feed and products. Thus, their MV can be elicited directly from market prices. In contrast, a public good is non-rival and non-excludable in consumption (Samuelson 1954). Thus, the consumption of the good by an individual does not preclude another from consuming the same good, and no one can be effectively excluded from using the good.

Animal welfare can be viewed as a public good (which is the animal product), since it is both non-excludable and non-rival in consumption (Lawrence et al. 2004).

For example, everyone is entitled to enjoy or benefit from good fish welfare, despite the fact that not everyone consumes farmed fish, hence this product is non-excludable. In addition, consumption of the fish does not reduce the amount of fish welfare as an ethical value, making it a non-rival good. In addition to being a public good, animal welfare is a characteristic of animal products, which is not or only partly traded and valued in the market today. Furthermore, studies show that market prices do not fully reflect the MV people place on animal welfare. The use of market prices as a sole measure of value will therefore underestimate the characteristic of animal welfare. McInerney (2004) emphasized that many of the animal welfare standards that society values may not be provided via market commodities, and that animal welfare, therefore, cannot be treated as a private good. So unless the improvement is at the upper end of our national welfare scale (i.e., no perceived benefits for a large proportion of the population), in terms of the textbook definitions, improved animal welfare must technically be considered as a public good (McInerney 2004).

Willingness to Pay for Improvements of Private and Public Goods

As will be shown in the following, consumer behavior, willingness to pay (WTP), and demand, all depend upon whether the good is defined as a private or public good such as animal welfare. We will first discuss WTP and demand of a private good; thereafter we will discuss supply and demand of animal welfare.

WTP for a quality improvement of private as well as public goods is defined as the maximum amount a person would be willing to pay, sacrifice or exchange for a good (Just et al. 1982). More formally, consumer WTP for a quality improvement is defined as the amount that makes the following equation hold:

$$V(\mathbf{p}, M - WTP, z_1) = V(\mathbf{p}, M, z_0) \quad (2)$$

where; V is the indirect utility function, \mathbf{p} is a vector of market prices, M is the amount of money available to the consumer (i.e., disposable income), and z_0 and z_1 are the quality levels of the good before and after the quality improvement, respectively. For private goods, the WTP for a quality improvement represents the price premium the consumers are willing to pay for the higher quality product.

The demand for a public good is derived differently from the demand of a private good. Non-rival consumption of public goods means that the total demand is obtained by aggregating the prices that all individuals are willing to pay for a given quantity. This is in contrast to a private good where the total market demand is obtained by aggregating the total quantities that all individuals are willing to buy at a given price.

McInerney (2004) discussed animal welfare and demand theory. He suggested that personal taste, preferences, and socio-economic factors (e.g., culture, education, media, campaigns, and social group) are the main factors determining the demand function for animal welfare. He argued that the market for welfare friendly livestock products is not greatly influenced by their price, because the ethical principles and personal values and feelings leading people to consider animal welfare provenance tend to be robust and not modified by prices. Instead, the market is influenced by people's education, awareness of welfare issues, and the normal evolution of perceptions and values in a generation. McInerney (2004) suggested a demand curve

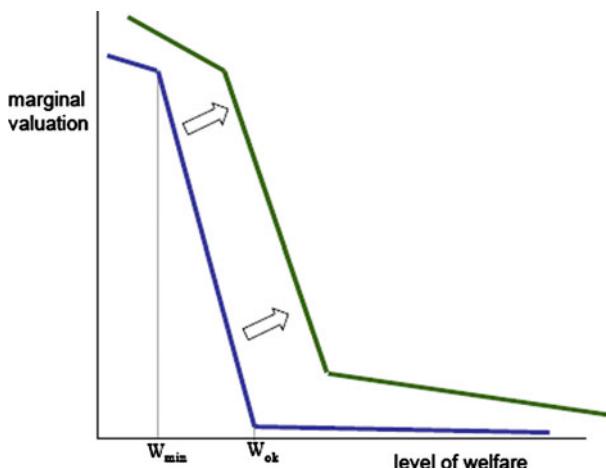


Fig. 1 Suggested nature of demand for animal welfare (McInerney 2004). Minimum point of animal welfare (W_{min}) and satisfactory point of animal welfare (W_{ok})

for animal welfare and illustrated the complexity of animal welfare as an attribute of animal products and how it should be handled in economic analyses. The demand curve for animal welfare by McInerney (2004) is illustrated in Fig. 1. The horizontal “quantity” axis represents increasing levels of animal welfare whereas the vertical “price” axis indicates the valuation people would attach to additional increments of welfare. When the level of animal welfare is low (below W_{min}), the value placed on improving welfare is high, because the conditions that animals face are of concern to most people. When the level of animal welfare exceeds this point of welfare, people have fewer concerns and consequently the marginal value they attach to improving animal welfare drops. When the level of what is considered as entirely satisfactory welfare conditions ($>W_{ok}$) is exceeded, little or no value is placed on a further increase in animal welfare. With an increase in people’s income or a change in attitude towards animal welfare in terms of caring more for the animals, the curve is expected to shift to the right. It was concluded (McInerney 2004), that even if markets for welfare-friendly products exist, data based on people’s WTP for these market goods does not capture the total value people place on animal welfare. In order to capture the missing non-market part of the total economic value, we also need to measure WTP for animal welfare in the general population; both for those who consume the animal product and those who do not.

Animal Welfare in Breeding Goals

Definition of Breeding Goal Values Related to Animal Welfare

The values used to weight the traits in the breeding goal (total economic value, TEV) can be divided into values of private goods driven by current market forces and values of public goods independent of current market forces (Fig. 2).

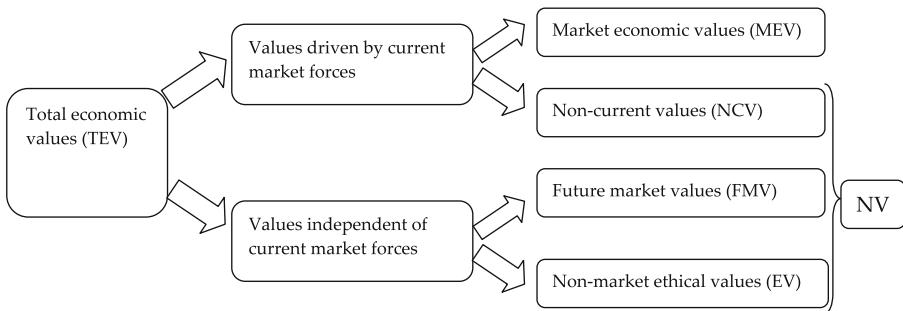


Fig. 2 Breeding goal values determining the breeding goal grouped according to whether they are driven by current market forces or independent of current market forces (NV = non-market value)

Values of animal welfare traits, which are directly linked to current market forces, will be included in the calculated MV because farmers and/or retailers get paid a higher price for the product from an animal that has had a higher level of animal welfare (in this paper we assume that the improved animal welfare is obtained through breeding for example by breeding animals with higher resistance to diseases). The MV correspond to economic values as given by Hazel (1943). MV can usually be defined using traditional methods such as profit equations and bio-economic models (e.g., Groen 1989; Nielsen et al. 2004).

When consumers are willing to pay for improved animal welfare, welfare is directly linked to market forces. In this paper we use the term non-current values (NCV) to denote a value, where it has been shown (e.g., through experimental market studies), that consumers are willing to pay extra for improved animal welfare but where an appropriate certification and labeling system is missing in the current market such that the products that differ with respect to grades of animal welfare cannot be differentiated from other products in the market. In order to quantify the value of, for example, improved welfare based on consumer WTP, products need to be labeled indicating that these differ in grades of animal welfare. Thus, certain animal welfare improvements become private goods with values and prices reflecting market demand and supply. Another example of a NCV could be legislation on grading of animal welfare related to diseases in farm animals. The government may use regulations on the frequency of the grade of, e.g., diseases in a population of animals, to obtain a disease frequency that is considered socially acceptable. Such a regulation makes it necessary to increase the value of the trait in the breeding goal related to animal welfare in order to breed animals with a higher resistance to diseases. This regulation would not necessarily change the market price of the products and in such a case the farmer will have to bear the cost associated with improved animal welfare. However, the governmental legislation may mimic a market force through fines, which makes an effective market incentive. In this case, such a value could be considered as a MV.

With respect to values that are independent of market forces, improved animal welfare can have a value even though there are no market forces directly or indirectly linked to it. A genetic change in traits may affect animal welfare, but

farmers or breeding companies are not paid more for the product even though the animal has better welfare through breeding. Here we can differentiate between two types of values; future market values (FMV) and ethical values (EV). In contrast to NCV, which people per definition is willing to pay for in the current market, FMV are not linked to market forces in the current market. An example of a FMV could be illustrated by a situation where consumers in the current market are unwilling to pay more for animals with a higher level of welfare such that the improvement in animal welfare will not be reflected in current prices of products. However, a breeding company may think that in the future their clients will demand animals with a higher level of animal welfare due to the anticipation that consumers or clients of the breeding companies are expected to be willing to pay more in the future. In Norway, for example, the dairy cattle breeding company GENO has for many years selected heavily on health and fertility traits. Today the breeding organization for Norwegian red dairy cattle are exporting semen to other countries such as the USA to be used for crossbreeding due to their genetic superiority in health and fertility traits (Hersleth 2005). Also, in New Zealand and Australia, a substantial number of sheep breeders include selection for animal resistance to internal parasites (based on recorded fecal egg counts of animals under artificial or natural challenges) as part of their breeding programs. In doing so, they sacrifice significant progress that could be made in production traits, and also incur significant recording costs. With effective drenching systems, the MV of internal parasite resistance is low. However, the calculated value (Amer 1999) is based on a sustainable drenching regime, whereby the expected build-up of genetic resistance by the parasite to anthelmintic treatment is expected to be minimal. The breeders who are aware of the risks of drench resistant parasites, and have ram-buying clients demanding such rams, choose to adopt an index including a value of resistance to parasites calculated in this way.

Although consumers may not directly demand animal welfare products today, we know that consumers are concerned about animal welfare. Hence, bad publicity regarding farm animal welfare may effectively introduce the market forces through drastically reduced consumer demand and prices of the animal product in question.

Regarding EV, improved animal welfare may have a value that will most likely never be linked to market forces. An example could be feet and leg problems in pigs (Kanis et al. 2005) and dairy cattle where a reduction in feet and leg problems in addition to a MV from saving costs, can have an EV. There may be an EV of reducing feet and leg problems because farmers, and other people involved, e.g., in handling the animals (at slaughter), may have less concerns about possible suffering of pigs and cows. This value refers more to the value people place upon animal welfare as citizens rather than as consumers (the difference between consumers and citizens was explained in the introduction), and may be affected by their willingness to support stricter governmental legislation and enforcement of better and more socially acceptable farm animal welfare (e.g., by citizens paying taxes and/or farmers paying fines for poor animal health and welfare). Finally, certain sections of the improvement in animal welfare traits will never be reflected in the market because market prices do not fully reflect the value people place on animal welfare (McInerney 2004; Olesen et al. 2006). We will elaborate further on this point later.

Accounting for Animal Welfare in the Breeding Goal

In order to include the value of improved animal welfare in the breeding goal beyond what can be directly estimated from the market economy (MV) we need to expand the breeding goal (see Eq. 1) by adding non-market values representing social, and ethical priorities to the breeding goal (Olesen et al. 2000). In that case the breeding goal (H) is a function of both MV and NV values for n traits:

$$H = MV_1 G_1 + MV_2 G_2 + \dots + MV_n G_n + NV_1 G_1 + NV_2 G_2 + \dots NV_n G_n, \quad (3)$$

where G_n is the additive genetic value for n traits, MV_n is the market economic value for n traits and NV_n is the non-market value for n traits.

The TEV placed on each trait is the sum of MV and NV. Note that the NV can be FMV, EV, or NCV according to the description of NV in “[Definition of Breeding Goal Values Related to Animal Welfare](#).[“](#) In addition, a given trait may have both a MV and a NV or only one of them. An example of a trait with both a MV and a NV is mastitis resistance in dairy cows, where the MV is derived based on the fact that increased resistance will yield fewer cases of mastitis with consequently lower costs of veterinarian treatments and lower losses due to discarded milk. In addition, mastitis resistance can have a NV due to improved welfare of the cow due to less suffering, stress, and discomfort from mastitis.

Genetic improvement as response to selection in a given trait (ΔG_i) can be valued by MV/or NV yielding market response and non-market response, respectively (Olesen et al. 2000):

$$\text{Market response: } MV_1 * \Delta G_1 + MV_2 * \Delta G_2 + \dots + MV_n * \Delta G_n$$

$$\text{Non-market response: } NV_1 * \Delta G_1 + NV_2 * \Delta G_2 + \dots + NV_n * \Delta G_n$$

The market response and non-market response express the revenue in monetary units one will get from the genetic improvement in the traits in the breeding goal. Total response to selection is then the sum of market response and non-market response.

Challenges when Incorporating Animal Welfare in the Breeding Goal

In this section we focus on the main challenges associated with considering animal welfare in the breeding goal.

Economic and Societal Values

The first main challenge of including welfare in the breeding goals comes from the fact that the value of improved animal welfare is not being fully reflected by prices and costs in the current market economy as was also addressed earlier in this review. In addition, consumers do not see animal welfare in food production mainly as their responsibility (Kjørstad 2005). People direct this responsibility to the government, producers, and retailers. Furthermore, consumers have limited knowledge about animal production and breeding (Quedrago 2003). This may also explain the gap

between attitudes and concerns about animal welfare and actual market behavior. This also indicates that market prices of labeled and animal friendly products do not reflect the total value people place on animal welfare.

Secondly, the views of many different stake holders (e.g., farmers, consumers, citizens, NGO's, and governmental authorities) need to be considered when defining breeding goals including welfare traits in order to include their concern about animal production (Gamborg and Sandøe 2005). When deriving MV these are usually derived at the farm level aiming at maximizing farmer profit. When including, e.g., consumer and citizens views on animal welfare concerns, the values need to be derived at a higher level of the production system (e.g., sector or national level). Deriving values at a higher level of the production system is complicated due to the increase in system size and often conflicting interests of the different actors (e.g., consumers and farmers).

Biological Factors

The fact that increased emphasis on traits related to animal welfare may reduce the response in production traits also bring in challenges of including welfare in the breeding goal. The reasons for the reduction in response are that:

- Including more traits in the breeding goal normally reduces the response per trait in the breeding goal.
- Relatively higher emphasis on non-production traits (such as resistance to diseases) affecting animal welfare will reduce the selection pressure on production traits.
- Non-production traits are often unfavorably correlated with production traits.

The above problems are further confounded by the fact that production traits usually have relatively high heritabilities and MV. This means that too much of a reduction in response in production traits due to an increased emphasis placed upon improving welfare traits, creates a large opportunity cost of lost selection response for traits returning MV to farmers and breeding companies. Therefore, improvement in traits related to animal welfare needs to be balanced with improvement in production and quality traits. Traits such as disease resistance, which are related to animal welfare, usually have low heritabilities, are difficult to measure, and in many cases are negatively correlated with production. Hence, they may deteriorate even when heavy selection emphasis is applied to them via weightings in a multiple trait breeding goal (Christensen 1998; Nielsen et al. 2006). Thus, including a trait related to animal welfare in the breeding goal is no guarantee for genetic improvement in the trait. Hence, appropriate weighting of traits through both MV and NV is needed to avoid unfavorable genetic changes.

Recently, breeding programs including social interactions have been suggested to overcome negative effects of selection (Bijma et al. 2007; Bergsma et al. 2008; Ellen et al. 2008). These social interactions between animals can influence selection response of traits related to animal welfare. In poultry, for example, social interactions and competition between hens can result in cannibalism and mortality (Ellen et al. 2008). The social effect (also called an associative effect) expresses

how the animal affects the phenotype of the other animals in the group (e.g., hens in a cage). By selecting the most productive animals, we may also select the most competitive or aggressive animals, which will have an impact of the performance of the other animals in the group (e.g., less growth and increased mortality). The social effect has a genetic component and by fitting this effect in statistical models to estimate breeding values (e.g., Bijma et al. 2007), it is possible to simultaneously improve production traits and traits affected by social interactions.

Methods to Derive Non-Market Values of Improved Welfare in Breeding Goals

Different methods have been suggested in order to derive NV in animal breeding goals (Olesen et al. 1999). The most relevant methods seem to be; desired gain indices, stated preference techniques such as the contingent valuation (CV) method (Mitchell and Carson 1989), and choice experiments (which are based on individual's WTP for a given product). In the following, we will review existing methods that can be used to derive NV of animal welfare and discuss their advantages and limitations. When the value of improved animal welfare is directly linked to current market forces, the MV can be derived using traditional methods such as profit equations or bio-economic models. The different methods and approaches were reviewed by for example Goddard (1998) and Groen et al. (1997), and will not be further considered here.

Methods Based on Stated Preferences Techniques

Consumer theory based on stated preference techniques such as the CV method (Mitchell and Carson 1989; Bateman and Willis 1999) or choice experiments (CE) based on stated choices (SC) or real choices (RC) (Carlsson and Martinsson 2001; Lusk and Schroeder 2004) can be used to estimate the value of animal welfare. In general, CV and CE are useful to estimate people's WTP for goods and products that are not traded in the market.

The CV method is based on surveys of individual's or household's WTP for a specified improvement of a good or service or willingness to accept compensation for a specified degradation of a good or service. Typically, a survey is performed in which people are asked how much money they would be willing to pay to obtain a marginal change in the quantity or quality of a public good. Thus, CV has become a popular method for valuing non-market public goods such as air and water quality (e.g., Desvouges et al. 1987), biodiversity (Garcia et al. 2009), cultural heritage (Navrud and Strand 2002), and public health (see review by Diener et al. 1998). Comprehensive accounts of the method can be found in Mitchell and Carson (1989) and Bateman and Willis (1999).

Applications for estimating relative MV in animal breeding goals are reported by Edel and Dempfle (2006), who estimated MV for conformation and performance of horses. They concluded that CV and SC may be valuable tools, but that incorporating the results in a breeding goal still remains a challenge. Another example of applying CV in animal breeding is the study by von Rohr et al. (1999), in which values for meat quality traits in pigs were derived. Meat quality experts

from slaughter and retail companies were asked how much they were willing to pay for specific products. Values for the meat quality traits were derived based on the answers from the interview. This is an example of deriving a NCV to be included in the breeding goal as the market prices for the different meat quality classes were not available from the market. Another example is the color of salmon fillet, which is currently not differentiated with respect to price in the market although people clearly have preferences for specific colors, and a WTP for this could be estimated by a choice experiment (Alfnæs et al. 2006; Steine et al. 2005). However, it may become technically possible to grade fillets with respect to color categories in future markets. Hence, fillet color is a trait that may have a higher market price in the future.

In a CE, respondents (e.g., consumers or individuals/households) are asked to view alternative descriptions of a good (e.g., meat or milk) with different attributes (description of the product e.g., level of welfare), and choose among products/goods at different levels of the attributes and at different prices. When respondents choose between the different alternatives in a given choice set, they implicitly make trade-offs between the levels of the attributes in the different alternatives in a given choice set.

There are two main types of CE; real choices (RC) and stated choices (SC). In a SC experiment consumers must choose between products with various quality traits in a series of shopping scenarios but there are no real economic incentives (Lusk and Schroeder 2004). A RC experiment is a non-hypothetical choice market experiment using posted prices and real economic incentives (Carlsson and Martinsson 2001; Cameron et al. 2002; Lusk and Schroeder 2004; Alfnæs et al. 2006). The participants are asked to make similar choices to the SC experiment, but the difference is that they need to buy and pay for the product.

Wurzinger et al. (2006) used a SC experiment in a Ugandan cattle breed to estimate farmers' relative preferences for cattle traits. An example of application of a RC experiment is the study by Olesen et al. (2010), where the aim was to estimate consumer's WTP for improved fish welfare. People were asked to make 10 choices between pairs of salmon filets that differed with respect to price and label of production system (i.e., conventional farm, organic production system or Freedom Food certified farm (certified by the animal welfare organization RSPCA)). Based on consumers' choices of salmon filet, the WTP extra for an organic or Freedom Food produced salmon filet was estimated to be 2 €/kg more than a conventional salmon filet. This example illustrates that if animal products can be labeled and marketed as being produced at a higher level of animal welfare, at least some of the value of improved welfare can be estimated. This also means that if it is possible to label products in the future, the value of improved animal welfare is no longer a NCV but a MV, because improved welfare of the fish will then be reflected in the market price.

Problems Associated with Methods Based on Stated Preferences Techniques A problem with accessing NV through WTP of labeled animal products is that in order to be able to use labeling in an animal breeding context, the labeling needs to represent individual traits (Nielsen and Amer 2007). For example in the study by Olesen et al. (2010), consumers were asked to choose between conventional and

organic produced salmon filet, which was not specified on a trait basis (e.g., improved fish welfare due to higher disease resistance). This makes it difficult to include the value of improved fish welfare in the breeding goal, because the value needs to be defined per unit of the trait. Another problem with labeling on an individual trait basis is that it may be difficult for the consumers to comprehend (Nielsen and Amer 2007). When considering surveys with consumers as respondents, performance levels of traits must be explained properly to the consumers (e.g., what mastitis is, and what the consequences are for the cow if infected with mastitis). Hence, appropriate and in depth CV studies or CE are probably needed for these types of analyses.

There are in general two prominent problems with hypothetical studies such as CV and SC. The first is known as the hypothetical bias, where participants overstate the amount they are willing to pay for an increase in quality of a private good (e.g., Cummings et al. 1995). This typically leads to positively biased WTP estimates. The second problem is the social desirability bias. Respondents like to appear different from what they are. This can occur consciously, because the respondents want to give the impression that they are socially responsible individuals, and may therefore embellish their answers according to how they think society would expect them to behave.

Due to the hypothetical bias in surveys, there has been a rising popularity of experimental markets in WTP studies (Shogren et al. 1994; List and Gallet 2001; List 2001) such as RC and second-price sealed-bid/Vickrey auctions, where the participants submit sealed bids for a product and the price is determined by the second-highest bid (Vickrey 1961; Shogren et al. 1994). The Vickrey auction is an incentive compatible method for eliciting WTP, i.e., it is in the participant's self interest to reveal their true preferences. However, the Vickrey auction is an unfamiliar market mechanism for most consumers. Consumers are used to markets where the seller posts the prices and they, as consumers, have to choose which products to buy.

WTP in the choice market is estimated with a logit or probit model from the choices made in the various scenarios. In this kind of experiment it is in the participant's own interest to choose the alternative they prefer in each scenario. Furthermore, the choice tasks in CE are relatively similar to the choices consumers face every day in grocery stores. This makes the RC experiment an incentive compatible method that can be used to estimate WTP for the characteristics of both private and public goods.

A general problem with CE is that in practice, with many traits in the breeding goal, it may be unrealistic to present the respondent with choices among all alternative combinations of traits and their levels. Therefore, it is important to generate and ask the survey questions in such a way that the maximum amount of information is collected from each respondent given other constraints such as survey cost, respondent's lack of knowledge, and the need to simplify the choices the respondents are asked to make. In addition, with many traits in the choice experiment, there is a risk that respondents may simplify the task by focusing only on the most important traits. Therefore, limits need to be placed on the number of traits that can be realistically examined (Carlsson and Martinsson 2001). Nielsen

and Amer (2007) showed how to simplify choices for respondents using a partial profile choice experiment. In contrast to a typical choice experiment, the choices in a partial profile design are simplified in the sense that only a subset of traits is presented to the respondent in each comparison. With a breeding goal of, for example six traits, respondents can be presented with only three or four of the six traits at a time.

Methods Based on Selection Index Theory

Derivation of values in the breeding goals is especially complicated when these are not directly linked to the current market (FMV and EV). Methods based on the selection index theory can be used to derive such values. These methods are, however, based on expert's preferences, as opposed to CV and CE, which are commonly based on consumer or citizens preferences.

Kanis et al. (2005) used so-called retrospective selection indices to define sustainable breeding goals for pig breeding programs. Their method is based on exploring the selection response surface for traits in the breeding goal by varying the TEV for traits in the breeding goal. Based on the selection response surface, breeding companies or farmers can then choose the TEV corresponding to the selection response in certain traits, which they find most sustainable or acceptable.

Nielsen et al. (2005, 2006) developed two methods to derive NV for traits related to sustainability such as mastitis resistance in dairy cattle. Both methods were based on selection index theory (Hazel 1943). A selection index is used to rank the different animals according to the breeding goals. In the first method, restricted indices were used to derive NV for mastitis and conception rate. The idea behind a restricted index is to restrict selection response in a given trait; e.g., to zero, meaning that no change in the trait is allowed. This can be used if there is an unfavorable genetic change in traits related to welfare (e.g., a decrease in resistance to diseases) due to negative correlations between production and diseases (see “[Biological Factors](#)”). It is then possible to estimate how much the TEV (the sum of NV and MV) should be on the disease trait in order to avoid a decline in this trait. The NV is then estimated as the TEV placed on the trait in order to avoid a decline in the trait minus the MV estimated purely from the market economy.

The second method by Nielsen et al. (2006) was also based on determining a certain response in a given trait. However, this method was based on how much farmers or breeding organizations are willing to lose in selection response for production traits in order to improve non-production traits.¹ Accepting a reduction in response in a production trait if needed, will allow for increased value and thus increased response in one or more traits related to animal welfare (in this case resistance to mastitis). The results for NV and selection response for resistance to mastitis are shown in Fig. 3 using an example with the traits milk yield, resistance to mastitis, conception rate, and still birth of calves. Assume a zero percent loss in

¹ The problem of reduction in selection response in production when there is increasing emphasis on non-production traits was explained in “[Challenges when Incorporating Animal Welfare in the Breeding Goal](#).”

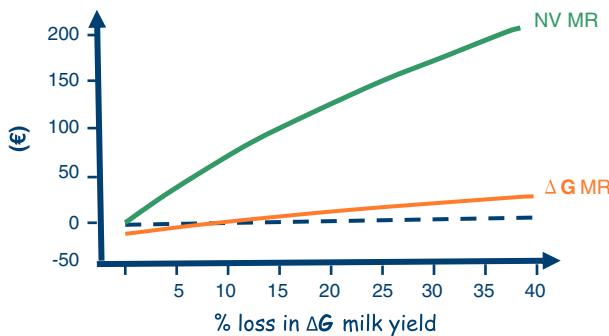


Fig. 3 Non-market values (NV MR) and selection response (ΔG MR) for resistance to mastitis in Euros as a function of percent loss in selection response in milk yield (mod. a. Nielsen et al. 2006)

selection response for milk yield corresponds to the situation where we only have MV of the traits in the breeding goal. The MV was calculated using a bio-economic model and was 0.28 €/kg for milk yield, 163 €/incidence for resistance to mastitis, 2.0 €/% for conception rate, and −39 €/% for stillborn calves (Nielsen 2004). As seen in Fig. 3, there is a negative response for mastitis resistance at this point, meaning that there will be more incidences of mastitis in the cows, even when selection for mastitis resistance is achieved by including the trait with MV in the breeding goal. However, if we then, for example, are willing to lose 10% in selection response in milk yield in order to improve mastitis resistance, we can increase the value on mastitis from 163 € (this is the MV) to 239 €. The non-market value is then the 239 € minus 163 € (the MV for mastitis resistance), which is 76 €. Therefore, by accepting a reduction in response in milk yield of 10% (compared with a situation where only MV is used in the breeding goal), we can increase the TEV on mastitis resistance and thereby increase response in mastitis resistance from −7.4 to 2.3 genetic standard deviations. Thus, we will then get a positive genetic response in resistance to mastitis and there will be fewer incidences of mastitis.

Advantages and Challenges Associated with Methods Based on Selection Index Theory One disadvantage with using a restricted index, where selection response in a given trait is restricted to zero change (Nielsen et al. 2005), is that it is assumed that the current level of resistance is ethically acceptable, which may not be the case. However, it is difficult to obtain an objective level for the desired change in a given trait. The method by Nielsen et al. (2006) gives a value of the response in a non-production trait relative to that in a production trait; which may be appealing to farmers, because production traits usually determine the farmer's main income. In addition, including the view of the farmer in deciding the level of trade-off between response in production versus response in non-production traits, may reduce the risk of the breeding goal being ignored. Moreover, it is possible to directly show the public how much response in production traits farmers or breeding organizations are willing to lose to improve non-production traits. The disadvantage is that the level of lost response in the production trait is chosen subjectively. However, it may be possible to obtain estimates for trade-offs between production traits and animal

welfare related traits from surveys with respect to people's willingness to support governmental legislation or subsidies in terms of being willing to pay e.g., higher income tax (Olesen et al. 2006).

In general, methods based on selection index theory have the advantage that they integrate the concept of deriving NV into classical breeding methodology, which may appeal to farmers and breeding companies. In addition, they are based on prediction of selection response for individual traits; which is an advantage, as it may be easier to communicate with breeding companies about selection response instead of TEV (Kanis et al. 2005), as a TEV in itself is not very informative about the actual emphasis on a given trait when selecting animals. However, selection responses for different traits are likely to be very unfamiliar terms for consumers and citizens.

Without subsidies, the incentives for breeding companies to produce animals with improved animal welfare will be low. Due to slow genetic progress for low-heritable traits related to animal welfare, breeding companies placing substantial selection emphasis on animal welfare traits may lose market share to competitors who have focused on short term improvement of more heritable production traits with higher MV. Therefore, it is likely that in many situations and for many reasons, farm animal breeders will place insufficient emphasis on animal welfare in their selection decisions. Under these circumstances, an outside party (e.g., a governmental body) may construct a selection objective that they feel is more appropriate to achieving improved animal welfare. Such an approach is advocated by Olesen et al. (2000), based on philosophical views of sustainability. However, this creates a serious risk of the selection objective being ignored. With an escalating awareness of corporate social responsibility, there seems to be a trend that breeding companies are increasingly focusing on their reputation and consumer concern as well as the impact these can have on long-term demand.

Supermarket chains may impose their buying power on suppliers, to make sure production practices are acceptable and to avoid bad publicity that particularly refers to the name of their chain. In the UK, for example, the chain Sainsbury's stopped selling eggs from caged hens before an EU-wide ban on battery produced eggs came into effect (Times online 2009). Although this is an example from animal production, one could imagine the same for a trait that could be changed by breeding (e.g., resistance to diseases or deformities). This effectively compensates for market failure, and creates a new market driver that is not explicitly linked to consumers paying more for differently labeled products (NCV). Breeding companies can apply methods based on desired gain indices (Nielsen et al. 2006) to increase genetic progress for traits related to animal welfare and, to some extent, keep up with competing companies on both the short and long term.

Translation of Stated Preference Techniques into Breeding Goal Values

Data from stated preference techniques are generally analyzed using random utility theory (Train 2003). Nielsen and Amer (2007) showed how data from a CE analyzed using conditional logit model could be transferred into MV and/or NV. The obtained solutions of regression coefficients ($\hat{\beta}$) can be used to estimate relative

values of traits in the breeding goal. TEV per unit genetic change in trait i' relative to a unit change in, e.g., trait $i = 1$ can be computed as:

$$\frac{(\beta_{i',j=n} - \beta_{i',j=1}) / (\alpha_{i',j=n} - \alpha_{i',j=1})}{(\beta_{i=1,j=n} - \beta_{i=1,j=1}) / (\alpha_{i=1,j=n} - \alpha_{i=1,j=1})}, \text{ for } i' \neq 1$$

where β_{ij} is the regression coefficient for trait i ($i = 1, \dots, m$) and level j ($j = 1, \dots, n$).

α_{ij} = absolute level of trait i at level j . The numerator is the value for trait i' expressed per trait unit by dividing by the change in the absolute level of the trait (the difference between the absolute level of the trait at level 3 ($\alpha_{i',j=3}$) and level 1 ($\alpha_{i',j=1}$)). The denominator is the value for trait 1, which ensures that the values of the traits are expressed relative to the value of trait 1.

One problem with the method described above is that it only yields MV/and or NV expressed in relative units (values expressed relative to each other). Values in monetary units are needed in order to calculate selection response in monetary units. These can possibly be obtained by including a price variable in the CE and then base TEV on WTP estimates. For a practical application of MV or NV in selection indices, these should be expressed per unit trait change, or alternatively per genetic standard deviation. To derive MV or NV, the survey questions then need to be developed on an individual traits basis (e.g., animal welfare related to individual traits) and they should include the performance level of the trait (Nielsen and Amer 2007). The challenge in practice is to present the traits in such a way that the respondents understand them, the unit they are measured in, and their different levels.

Implications for Future Work

As is clear from this review, including animal welfare in breeding goals requires a multidisciplinary approach and requires participation at different levels.

The main challenges associated with including welfare in the breeding goal in our opinion are:

1. How to derive the EV of animal welfare that cannot be captured by measuring people's WTP for improved animal welfare as in an experimental market study.
2. How to translate data from stated preference techniques into values to be applied in breeding goals and selection indices.

Firstly, if animal products can be labeled and marketed as having been produced with better animal welfare, and if people are willing to pay a higher price for a product with better animal welfare, some of the value of animal welfare can be captured by the higher market price. This implies that characteristics that do not have a market value today may get one in a future market (for example, this may correspond to a FVM or possible NCV). MV of the characteristics of labeled products such as "Freedom Food," which is an international label with stricter

animal welfare production criteria than the standard criteria in Europe, can then be obtained by experimental markets as presented by Steine et al. (2005), Alfnes et al. (2006), and Olesen et al. (2010). These can be regarded as a minimum value of improved animal welfare connected to the product. Possible EV will add to these “minimum” values.

The challenge is to derive these EV, since studies indicate that consumers do not see animal welfare in food production to be their responsibility (Kjørstad 2005). The same author found that in Sweden, the Netherlands, the UK, and Norway, most people direct this responsibility to the government and secondarily to the producers and the retailers. This may explain why there is a gap between attitudes and concerns between animal welfare, on the one hand, and eating practices, on the other. This is especially noticeable in Norway and Sweden (Kjørstad 2005). Furthermore, this demonstrates that market prices of labeled and animal-friendly products do not reflect the TEV people place on animal welfare. For capturing more of people's and society's total valuation of animal welfare, other studies of individual's attitudes and WTP for alternative production systems and breeding programs are needed. In order to derive these EV of animal welfare it is necessary to engage with different participants in animal breeding and production, such as breeding organizations, food authorities, and animal welfare organizations. Here, discussion is needed on how animal welfare should be emphasized and given priority, relative to higher expected profits for the farmer or lower prices to the consumer.

Furthermore, governmental restrictions on genetic changes of animal welfare traits may be applied according to social non-acceptance of compromised animal welfare. This could, e.g., be through governmental restrictions and supervision on the non-use of breeding animals with low breeding merit for welfare related traits and on a multiple trait breeding goal that assures genetic progress or status quo for welfare-related traits. The latter may result in lower expected profits for farmers than under the current rules and regulations. This could possibly be used to determine an acceptable level of loss in selection response for production traits (loss in market response) in order to improve traits related to animal welfare.

With respect to how to formulate questions in stated preference techniques, the problem is that consumers may in general have poor knowledge about breeding methods. In a study of Quedrado (2003), which was part of the SEFABAR (Sustainable European Farm Animal Breeding and Reproduction) project; consumers were concerned about the impact of breeding methods on their food items, although participants rarely mentioned breeding as a food-related concern. The concerns raised by survey respondents included safety, health effects, food quality, animal welfare, and environmental impacts. To some degree, it seemed from the survey that production, breeding, and reproduction issues were combined. The survey was based on so-called focus group interviews in the UK and France (Quedrado 2003). Thus, before using SP techniques it is important to have sessions with focus groups in order to gain knowledge about consumers/citizens knowledge regarding breeding and animal welfare.

In addition, as stated in “[Methods Based on Stated Preferences Techniques](#),” for a practical implementation of TEV it is important that these can be connected to the

level of the trait. However, it may be very difficult for consumers/citizens in a survey to relate to changes in individual animal traits. It may be possible to merge a group of traits into one trait (e.g., general resistance to diseases) in the survey and then estimate NV for individual disease traits based on disease frequencies in a given population. Alternatively, this weighting of the different individual traits could be done by weighting by the degree of suffering associated with each disease as judged by farmers and/or veterinarians.

In order to include the value of animal welfare in the breeding goal we suggest to:

1. Use profit equations to derive market values (MV).
2. Define traits in the breeding goal related to animal welfare.
3. Use stated preference techniques to obtain NCV or FMV from, e.g., labeled products, and for obtaining EV for specific welfare traits with non-market value. RC experiments can be applied for products that may be marketed, whereas SC experiments and CV studies may be applied to EV.
4. Sessions with focus groups should be held to obtain knowledge about consumer and/or citizen knowledge about animal welfare and breeding; and proper questions for the surveys and possible experiments should then be formulated.
5. Market, non-market, and total selection response should be predicted for the traits in the breeding goal.
6. Methods based on selection index theory should be applied to determine trade-offs in production traits versus traits related to animal welfare.
7. Finally, the results should be discussed with stakeholders in animal breeding and food production.

Conclusion

In this paper we have reviewed the current status of available methods and associated challenges for including animal welfare in breeding goals. Even though these methods provide the tools necessary for considering welfare issues in the breeding goal, there are few practical applications of the methods. Such applications require decisions to be made about whether farmers (and/or the consumer) should bear all the cost of improving traits related to animal welfare or whether society should cover some of the losses, e.g., through regulations/subsidized breeding since animal welfare also has an EV. Definition of breeding goals including values related to animal welfare requires a multidisciplinary approach. This includes a combination of different methods such as profit equations, SP techniques and selection index theory, and, in addition, a participatory approach involving different stakeholders such as breeding organizations, food authorities, farmers, and animal welfare organizations.

Acknowledgments The study was financed by The Research Council of Norway, grant no. 143045/140 (Future Animal Breeding Goals) and grant no. 190457/S40 (Breeding for Improved Fish Welfare in Aquaculture). We would like to thank the anonymous reviewers of a previous version of this paper for their helpful comments.

References

- Alfnes, F., Guttormsen, A. G., Steine, G., & Kolstad, K. (2006). Consumers' willingness to pay for the color of salmon: A choice experiment with real economic incentives. *American Journal of Agricultural Economics*, 88, 1050–1061.
- Amer, P. R. (1999). Economic accounting of numbers of expressions and delays in sheep genetic improvement. *New Zealand Journal of Agricultural Research*, 42, 325–336.
- Bateman, I., & Willis, K. G. (1999). *Valuing environmental preferences. Theory and practice of the contingent valuation method in the US, EU, and developing countries*. Oxford, UK and New York, USA: Oxford University Press.
- Bergsma, R., Kanis, E., Knol, E. F., & Bijma, P. (2008). The contribution of social effects to heritable variation in finishing traits of domestic pigs (*Sus scrofa*). *Genetics*, 178, 1559–1570.
- Bijma, P., Muir, W. M., & Arendonk, J. A. M. (2007). Multilevel selection 1: Quantitative genetics of inheritance and response to selection. *Genetics*, 175, 277–288.
- Brascamp, E. W., Smith, C., & Guy, D. R. (1985). Derivation of economic weights from profit equations. *Animal Production*, 40, 175–180.
- Broom, D. M. (1991). Animal welfare: Concepts and measurement. *Journal of Animal Science*, 69, 4167–4175.
- Cameron, T., Poe, G., Either, R., & Schulze, W. (2002). Alternative nonmarket value-elicitation methods: Are revealed and stated preferences the same? *Journal of Environmental Economics and Management*, 44, 391–421.
- Carlsson, F., & Martinsson, P. (2001). Do hypothetical and actual marginal willingness to pay differ in choice experiments? *Journal of Environmental Economics and Management*, 41, 179–192.
- Christensen, L. G. (1998). Possibilities for genetic improvement of disease resistance, functional traits and animal welfare. *Acta Agriculturae Scandinavica. Section A, Animal Science, Supplement*, 28, 77–89.
- Cummings, R. G., Harrison, G. W., & Rustrom, E. E. (1995). Homegrown value and hypothetical surveys: Is the dichotomous choice approach incentive-compatible? *American Economic Review*, 85, 260–266.
- Dekkers, J. C. M., & Gibson, J. P. (1998). Applying breeding objectives to dairy cattle improvements. *Journal of Dairy Science*, 81, 9–35.
- Desvouges, W. H., Smith, V. K., & Fisher, A. (1987). Option price estimates for water quality improvements: A contingent valuation study for the monongahela river. *Journal of Environmental Economics and Management*, 14, 248–267.
- Diener, A., O'Brien, B., & Gafni, A. (1998). Health care contingent valuation studies: A review and classification of the literature. *Health Economics*, 7, 313–326.
- Edel, C., Dempfle, L. (2006). The use of the contingent valuation method to define breeding objectives: Experiences from two studies on horse breeds. In Proc. 8th World Congress on Genetics Applied to Livestock Production. 13–18 Aug 2006, Belo Horizonte, Brazil. CD-ROM: 31_368-1668.pdf. pp. 4.
- Ellen, E. D., Visscher, J., van Arendonk, J. A. M., & Bijma, P. (2008). Survival of laying hens: Genetic parameters for direct and associative effects in three purebred layer lines. *Poultry Science*, 87, 233–239.
- Fraser, D. (2003). Assessing animal welfare at the farm and group level: The interplay of science and values. *Animal Welfare*, 12, 433–443.
- Gamborg, C., & Sandøe, P. (2005). Sustainability in farm animal breeding: A review. *Livestock Production Science*, 92, 221–231.
- Garcia, S., Harou, P., Montagné, C., & Stenger, A. (2009). Models for sample selection bias in contingent valuation: Application to forest biodiversity. *Journal of Forest Economics*, 15, 59–78.
- Gibson, J. P., & Wilton, J. W. (1998). Defining multiple-trait objectives for sustainable genetic improvement. *Journal of Animal Science*, 76, 2303–2307.
- Goddard, M. E. (1998). Consensus and debate in the definition of breeding objectives. *Journal of Dairy Science*, 81, 6–18.
- Groen, A. F. (1989). Economic values in cattle breeding. I. Influences of production circumstances in situations without output limitations. *Livestock Production Science*, 22, 1–16.
- Groen, A. F., Steine, T., Colleau, J., Pedersen, J., Pribyl, J., & Reinsch, N. (1997). Economic values in dairy cattle breeding with special reference to functional traits. Report of an EAAP-working groups. *Livestock Production Science*, 49, 1–21.

- Hazel, L. N. (1943). The genetic basis for constructing selection indexes. *Genetics*, 28, 476–490.
- Heringstad, B., Klemetsdal, G., & Ruane, J. M. (2000). Selection for mastitis resistance in dairy cattle: A review with focus on the situation in the Nordic countries. *Livestock Production Science*, 64, 95–106.
- Hersleth, E. (2005). NRF Norwegian red breed—a source for crossbreeding. Paper presented at the Red cow symposium: Breeding to meet the changing economics of the dairy industry. South Afrika.
- Horgan, R., & Gavinelli, A. (2006). The expanding role of animal welfare within EU legislation and beyond. *Livestock Production Science*, 103, 303–307.
- Just, R., Hueth, D., Schmitz, A. (1982). *Applied welfare economics and public policy*. New York, US.
- Kanis, E., de Greef, K. H., Hiemstra, A., & van Arendonk, J. A. M. (2005). Breeding for societally important traits in pigs. *Journal of Animal Science*, 83, 948–957.
- Kanis, E., van den Belt, H., Groen, A. F., Schakel, J., & de Greef, K. H. (2004). Breeding for improved welfare in pigs: A conceptual framework and its use in practise. *Animal Science*, 78, 315–329.
- Kjærnes, U., Lavik, R., Kjørstad, I. (2005). Animal friendliness and food consumption practices. Preliminary results from population surveys in 7 countries. In A. Butterworth (Ed.), *Science and society improving animal welfare. Welfare Quality conference proceedings*, 17–18 Nov 2005, Brussels, Belgium, pp. 26–33.
- Kjørstad, I. (2005). Welfare quality. WP1.1 consumers literature review country report Norway. SIFO, National Institute for Consumer Research, P.O.Box 4682 Nydalen, N-0405 Oslo Norway. pp. 60.
- Lawrence, A. B., Conington, J., & Simm, G. (2004). Breeding for animal welfare: Practical and theoretical advantages of multi-trait selection. *Animal Welfare*, 13, 191–196.
- List, J. A. (2001). Do explicit warnings eliminate the hypothetical bias in elicitation procedures? Evidence from field auctions for sportcards. *American Economic Review*, 91, 1498–1507.
- List, J. A., & Gallet, G. A. (2001). What experimental protocol influence disparities between actual and hypothetical stated values? *Environmental Resource Economics*, 20, 241–254.
- Lusk, J. L., & Schroeder, T. C. (2004). Are choice experiments incentive compatible? A test with quality differentiated beef steaks. *American Journal of Agricultural Economics*, 86, 467–482.
- McInerney, J. (2004). *Animal welfare, economics and policy*. Report on a study undertaken for the Farm & Animal Health Economics Division of Defra. pp. 68.
- Mitchell, R., & Carson, R. (1989). *Using surveys to value public goods: The contingent valuation method*. Washington, DC: Resources for the Future.
- Navrud, S., Strand, J. (2002). Social cost and benefits of preserving and restoring the Nidaros Cathedral. In S. Navrud & R. C. Ready (Eds.) *Valuing cultural heritage: Applying environmental valuation techniques to historic buildings, moments and artifacts*. Northampton: Edward Elgar Publishing.
- Nielsen, H. M. (2004). Economic values for production and functional traits in dairy cattle breeding goals derived by stochastic simulation. Ph.D. Thesis, Royal Veterinary and Agricultural University, Copenhagen, Denmark.
- Nielsen, H. M., & Amer, P. R. (2007). An approach to derive economic weights in breeding objectives using partial profile choice experiments. *Animal*, 1, 1254–1262.
- Nielsen, H. M., Christensen, L. G., & Groen, A. F. (2005). Derivation of sustainable breeding goals for dairy cattle using selection index theory. *Journal of Dairy Science*, 88, 1882–1890.
- Nielsen, H. M., Christensen, L. G., & Ødegård, J. (2006). A method to define breeding goals for sustainable dairy cattle production. *Journal of Dairy Science*, 89, 3615–3625.
- Nielsen, H. M., Groen, A. F., Pedersen, J., & Christensen, L. G. (2004). Stochastic simulation of economic values and their standard deviations for production and functional traits in dairy cattle under current and future Danish production circumstances. *Acta Agriculturae Scandinavica*, 54, 113–126.
- Olesen, I., Alfnæs, F., Røra, M. B., & Kolstad, K. (2010). Eliciting consumers' willingness to pay for organic and welfare-labelled salmon in a non-hypothetical choice experiment. *Livestock Science*, 127, 218–226.
- Olesen, I., Gjerde, B., & Groen, A. F. (1999). Methodology for deriving non-market trait values in animal breeding goals for sustainable production systems. In Proceedings International Workshop on EU Concerted Action on Genetic Improvement of Functional Traits in Cattle (GIFT); Breeding Goals and Selection Schemes, Wageningen, The Netherlands, 7–9th Nov 1999. Interbull Bulletin No. 23 (1999), pp. 13–21.
- Olesen, I., Groen, A. F., & Gjerde, B. (2000). Definition of animal breeding goals for sustainable production systems. *Journal of Animal Science*, 78, 570–582.

- Olesen, I., Navrud, S. & Kolstad, K. (2006). Economic values of animal welfare in breeding goals. In Proc. 8th World Cong. Genet. Appl. Livest. Prod. 13–18 Aug 2006, Belo Horizonte, Brazil. CD-ROM: 31_538-1710.pdf. pp. 5.
- Philipsson, J., & Lindhé, B. (2003). Experiences of including reproduction and health traits in Scandinavian dairy cattle breeding programmes. *Livestock Production Science*, 83, 99–112.
- Quedago, A. P. (2003). Symbolic goods in the market place. Public perceptions of farm animal breeding and reproduction in France and United Kingdom. In A. E. Liinamo & A. M. Neeteson-van Nieuwenhoven (Eds.) Pages 36–46 in SEFABAR (Sustainable European Farm Animal Breeding and Reproduction), Final Workshop, Rome, Italy.
- Randall, A. (1987). *Resource economics* (p. 434). New York: John Wiley & Son.
- Rauw, W. M., Kanis, E., Noordhuizen, E. N., & Grommers, F. J. (1998). Undesirable side effects of selection for high production efficiency in farm animals: A review. *Livestock Production Science*, 56, 13–33.
- Sagoff, M. (1988). *The economy of the earth*. Cambridge: Cambridge University Press.
- Samuelson, P. A. (1954). The pure theory of public expenditure. *Review of Economics and Statistics*, 36, 387–389.
- Sandøe, P., Christiansen, S. B., & Appleby, M. C. (2003). Farm animal welfare: The interaction of ethical questions and animal welfare science. *Animal Welfare*, 12, 469–478.
- Sandøe, P., Nielsen, B. L., Christensen, L. G., & Sørensen, P. (1999). Staying good while playing god—the ethics of breeding farm animals. *Animal Welfare*, 8, 313–328.
- Shogren, J. F., Shin, S. Y., Hayes, D. J., & Kliebenstein, J. B. (1994). Resolving differences in willingness to pay and willingness to accept. *American Economic Review*, 84, 255–270.
- Star, L., Ellen, E. D., Uitdehaag, K., & Brom, F. W. A. (2008). A plea to implement robustness into a breeding goal: Poultry as an example. *Journal of Agricultural and Environmental Ethics*, 21, 109–125.
- Steine, G., Alfnes, F., & Rørå, M. B. (2005). The effect of color on consumer WTP for farmed salmon. *Marine Resource Economics*, 20, 211–219.
- Timesonline. (2009). http://www.timesonline.co.uk/tol/life_and_style/food_and_drink/article5535331.ece. Accessed 10 April 2010.
- Torp Donner, H., & Juga, J. (1997). Sustainability—a challenge to animal breeding. *Agricultural Food Science Finland*, 6, 229–239.
- Train, K. E. (2003). *Discrete choice methods with simulation*. Cambridge, UK: Cambridge University Press.
- Van Dorp, T. E., Dekkers, J. C. M., Martin, S. W., & Noordhuizen, J. P. T. M. (1998). Genetic parameters of health disorders, and relationships with 305-Day milk yield and conformation traits of registered Holstein cows. *Journal of Dairy Science*, 81, 2264–2270.
- Vickrey, W. (1961). Counterspeculation, auctions, and competitive sealed bids. *Journal of Finance*, 16, 8–37.
- von Rohr, P., Hofer, A., & Kunzi, N. (1999). Economic values for meat quality traits in pigs. *Journal of Animal Science*, 77, 2633–2640.
- Wurzinger, M., Ndumu, D., Baumung, R., Drucker, A. G., Okeyo, A. M., Semambo, D. K., Sölkner, J. (2006). Assessing stated preferences through the use of choice experiments: Valuing (re)production versus aesthetics in the breeding goals of Ugandan Ankole cattle breeders. In Proc. 8th World Cong. Genet. Appl. Livest. Prod. 13–18 Aug 2006, Belo Horizonte, Brazil. CD-ROM: 31_439-608.pdf. pp. 4.