

3501-2011

Genetische Grundlagen der Pflanzenzüchtung

B. Sc. Agrarwissenschaften 5. Semester

Wintersemester 2024/2025

Version: 15. Oktober 2024

Prof. Dr. Karl Schmid

Fachgebiet Nutzpflanzenbiodiversität und
Züchtungsinformatik (350b)

Institut für Pflanzenzüchtung, Saatgutforschung und
Populationsgenetik



**UNIVERSITÄT
HOHENHEIM**

3501-211

Genetische Grundlagen der Pflanzenzüchtung

B.Sc. Agrarwissenschaften 5. Semester

1 / 9

Contact information

Prof. Dr. Karl Schmid



- Research area: Crop biodiversity and breeding informatics
- Fruwirthstrasse 21
- Telephone: 0711/459 23487
- Email: karl.schmid@uni-hohenheim.de
- Web: <http://evoplant.uni-hohenheim.de>
- Office hours: After appointment by email

2 / 9

Pflanzenzüchtung und SaatgutBunde (3501-210)

WiSe 2023/2024

Cooperator: Prof. Wissel - Institute 203 for breeding, Seed Science and Population Genetics

Prof. Schmid (070) 3501-211 (Deutsche Grundlagen der Pflanzenzüchtung)

Prof. Wissel (0706) 3501-212 (Allgemeine Pflanzenzüchtung)

Prof. Klose (0706) 3501-213 (Saatguttechnik und Saatgutproduktion)

Prof. Dr. Schmid (0706) 3501-214 (Qualitative Genetik, Variationslehre zum Pflanzengen)

Do 17.10.2023 14:00 - 16:10 **Qualitative Genetik** Prof. Schmid H 4.1

Do 18.10.2023 14:00 - 16:10 **Qualitative Genetik und Quantitative Genetik von Mutterpflanzen** Prof. Schmid H 4.2

Do 24.10.2023 14:00 - 16:00 **Quantitative Vererbungsgruppen von Mutterpflanzen** Prof. Schmid H 4.1

Do 26.10.2023 14:00 - 16:00 **Das Genom von Pflanzen: Struktur und Variation** Dr. A. Stoebe H 4.2

Do 02.11.2023 14:00 - 16:00 **Reziproker Kreuzungstest: Quantitative Vererbungsgruppen** Prof. Schmid H 4.2

Do 07.11.2023 14:00 - 16:00 **Quantitative Genetik: Variationslehre zum Pflanzengen** Prof. Schmid H 4.1

Do 08.11.2023 14:00 - 16:00 **Rekombination in plant breeding** Prof. Wissel H 4.2

Do 14.11.2023 14:00 - 16:00 **Rekombination in plant breeding: crossing strategies** Prof. Wissel H 4.2

Do 16.11.2023 14:00 - 16:00 **Breeding open-pollinated varieties** Prof. Wissel H 4.2

Do 21.11.2023 14:00 - 16:00 **Breeding closed varieties** Prof. Dr. H. Klose H 4.2

Do 23.11.2023 14:00 - 16:10 **Breeding the varieties** Prof. Dr. H. Klose H 4.2

Do 07.12.2023 14:00 - 16:00 **Recombinant breeding: backcross breeding** Prof. Wissel H 4.2

Do 08.11.2023 14:00 - 16:00 **Breeding hybrid varieties I** Prof. Wissel H 4.2

Do 09.11.2023 14:00 - 16:00 **Breeding hybrid varieties II** Prof. Wissel H 4.2

Do 13.12.2023 14:00 - 16:00 **Comparison of breeding categories** Prof. Wissel H 4.2

Do 14.12.2023 14:00 - 16:00 **The backcross method** Prof. Wissel H 4.2

Do 19.12.2023 14:00 - 16:00 **Repetitorium and Questions I** Prof. Wissel H 4.2

Do 20.12.2023 14:00 - 16:00 **Repetitorium and Questions II** Prof. Wissel H 4.2

Do 06.01.2024 14:00 - 16:00 **Argonautische Deltapflanze: mendelige Arten** Prof. Klose H 4.2

Do 11.01.2024 14:00 - 16:00 **Argonautische Deltapflanze: dihybride Arten** Prof. Klose H 4.2

Do 12.01.2024 14:00 - 16:00 **Argonautische Deltapflanze: tetrahybride Arten** Prof. Klose H 4.2

Do 18.01.2024 14:00 - 16:00 **Feststellung und Herleitung** Prof. Klose H 4.2

Do 19.01.2024 14:00 - 16:00 **Breeding self pollinating** Prof. Klose H 4.2

Do 23.01.2024 14:00 - 16:00 **Breeding products of heterothallic species** PD Dr. M. Nagel H 4.2

Do 25.01.2024 14:00 - 16:00 **Breeding products of hermaphroditic species** PD Dr. M. Nagel H 4.2

Do 30.01.2024 14:00 - 16:00 **Biotechnology in plant breeding** PD Dr. M. Nagel H 4.2

Do 07.02.2024 14:00 - 16:00 **Saatgutprüfung und Gerücksicht** PD Dr. M. Nagel H 4.2

Klausur 1

Klausur 2

3 / 9

Why are the slides in English language?

English is the **lingua franca** of science.

It is important to learn (and master) the english language at an early stage

The reading of scientific literature, which is usually written in English, is greatly facilitated with a good knowledge of English.

4 / 9

Motivation for the course



Key questions:

- What is the substance of inheritance?
- How can inheritance be manipulated?
- What are the rules of inheritance?
- How is genetic variation inherited?
- How is phenotypic variation inherited?

5 / 9

Exam preparation

- Our part of the exam consists of open questions (No multiple choice)
- Answer review questions
- Read the literature provided

6 / 9

General reading (for the module)

- Heiko Becker: Pflanzenzüchtung (2019) 3. Auflage, Ulmer UTB
- George Acquaah: Principles of Plant Genetics and Breeding (2012) 2nd edition, Wiley-Blackwell [Acquaah, George, 2012]
- Griffiths et al.: Introduction to Genetic Analysis (2019) 12th edition, Freeman

7 / 9

Übersicht über Vorlesungen

- Zell- und Molekularbiologie
- Biotechnologie und Gentechnik
- Genetische Vererbungsregeln von Mendel
- Molekulare Marker und genetische Karten
- Populationsgenetik I: Hardy-Weinberg-Gleichgewicht
- Populationsgenetik II: Genetische Drift und Selektion
- Quantitative Genetik: Vom Genotyp zum Phänotyp

8 / 9

References i

-  Acquaah, George (2012). *Principles of Plant Genetics and Breeding*. Wiley, 2nd edition.

9 / 9



Cell and Molecular Biology

3501-211 Genetische Grundlagen der Pflanzenzüchtung

Prof. Karl Schmid
WS 2023/2024

Institute of Plant Breeding, Seed Science and Population Genetics
University of Hohenheim

1 / 51

Overview

The eucaryotic cell

The genome of plants

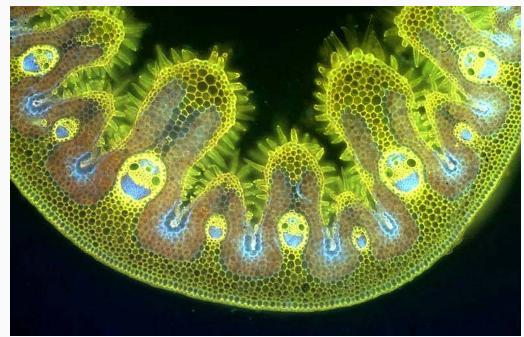
Organization of plant genomes

Replication of DNA

The structure and expression of genes

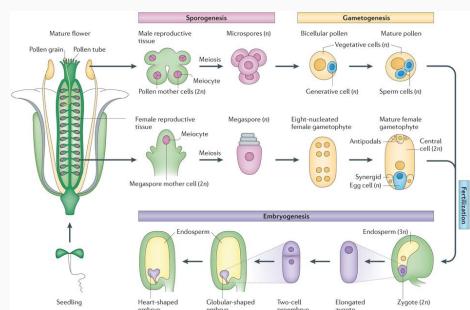
2 / 51

Plants are made up of cells and cells are made up of molecules



3 / 51

Why are cell and molecular biology relevant for genetics?



[Kawashima and Berger, 2014]

4 / 51

Genetics deals with the inheritance of genes and traits.

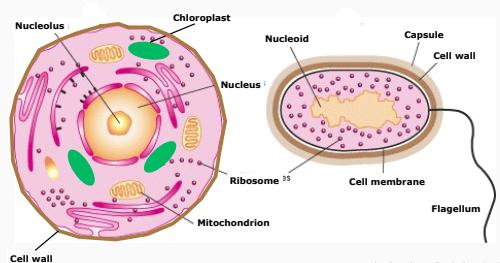
Inheritance is carried out by different types of molecules and occurs in germ cells. It is a highly integrated and very complicated process. The relevance of cell and molecular biology for plant breeding can be shown with this figure.

It shows the process of reproduction, which is the basis of inheritance. You can see that many different types of cells are involved. In addition, many different types of molecules are involved in the interaction between male and female components of plant reproduction.

Since plant breeding is dependent on successful reproduction both the understanding on a fundamental level and the manipulation of cellular and molecular processes is essential.

The same can be said for other aspects of plant breeding such as the resistance to drought or other stress types, the interactions with pathogens, the yield and quality of the product.

The structure of a cell



www.ncbi.nlm.nih.gov/books/NBK21120

5 / 51

On this slide there are two cells.

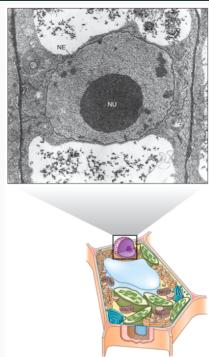
The left is a eukaryotic cell of a plant, and on the left a prokaryotic cell of a bacterium. One can see important differences between both cell types.

The key difference is that the plant cell has a plant nucleus that contains the DNA, whereas the prokaryotic cell has a nucleoid structure that contains the bacterial DNA.

Another difference is the size of the cell (eukaryotic plant cells can vary greatly in size), and additional structures, which are the plant organelles.

Important plant organelles are chloroplasts and mitochondria. They have their own genome and are also inherited differently.

The plant nucleus



6 / 51

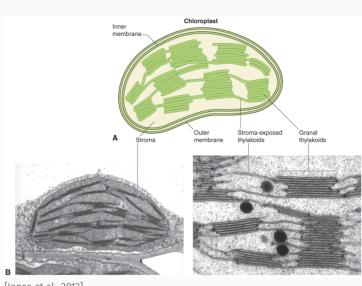
The nucleous is a key organelle of eukaryotes.

It contains the nucleoplasm, in which the genetic material, or chromatin, is found.

The nuclear envelope is made up of phospholipid bilayers (a bilayer is the basic unit of a membrane) with many pores, that allow an exchange of molecules between nucleoplasm and cytoplasm.

The nucleus contains the chromosomes, which consist of DNA-protein complexes, or chromatin.

Chloroplast



- Circular double-stranded DNA
- Multiple circular DNA molecules
- Photosynthesis, hydrolysis of water, oxygen production

7 / 51

There are two other important organelles in the plant cell.

The chloroplast consists of internally stacked membranes, the thylakoids.

Its main function is the conversion of light, water and carbon dioxide into carbohydrates.

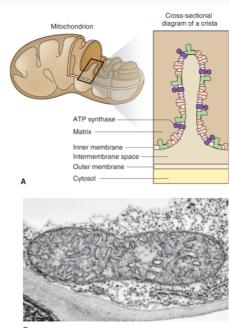
Chloroplasts harbor their own genome in forms of multiple circular DNA molecules.

Not all genes required for chloroplast functions are located on the chloroplast genome. For this reason, many proteins must be imported into the chloroplast for its full function.

Chloroplasts divide by fission.

They are only maternally inherited with the cytoplasm.

Mitochondria



[Jones et al., 2012]

- Circular double-stranded DNA
- Multiple circular DNA molecules
- Respiration, ATP, carbon dioxide production

8 / 51

Mitochondria are the powerhouses of the cell and occur in almost all plant cells.

In mitochondria different processes take place: The aerobic respiration, which converts carbohydrates into energy and carbon dioxide, the gluconeogenesis, which is the breakdown of lipids (fat) into carbohydrates, and the photrespiration, which is a light-stimulated process that competes with photosynthesis because it consumes oxygen and produces carbon dioxide.

Like chloroplast, mitochondria have their own genome and divide by fission. They are maternally inherited.

Mitochondria have two membrane systems, an outer and an inner membrane. The inner membrane, or mitochondrial matrix, is densely populated with enzymes located on cristae, that perform the respiratory chain which leads to the production of energy.

The eukaryotic cell

The genome of plants

Organization of plant genomes

Replication of DNA

The structure and expression of genes

9 / 51

Organisation of plant genomes

A **genome** is the total DNA content of a plant cell.

Plant genomes are organized into three compartments:

- Nucleus: **Nuclear genome**
- Chloroplast and other plastides: **Plastid genome**
- Mitochondrion: **Mitochondrial genome**

A **gene** is a functional unit in a genome

10 / 51

How are genome sizes measured?

- **bp** one pair of nucleotide bases
- **kb** 1 kilobase (kilobase pairs) = 1,000 bp
- 1 megabase (megabase pairs) = 1,000,000 bp
- **Gb** 1 gigabase (gigabase pairs) = 1,000,000,000 bp

Typical sizes of organelle genomes:

- Chloroplast: 120 - 160 kb
- Mitochondria: 200 kb - 2600 kb

Organelle genomes originated from prokaryotic **endosymbionts**.

11 / 51

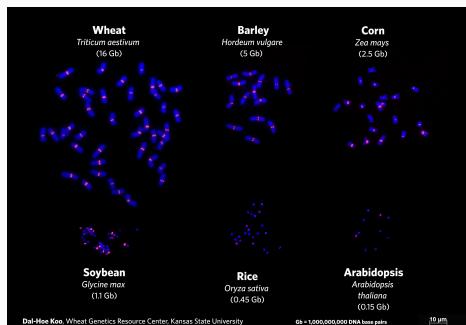
Mitochondrial genomes show much larger size variation than plastid genomes.

The origin of organelles derives from the merger of an early eukaryote with a photosynthetic cyanobacterium, which subsequently evolved into the chloroplast.

The mitochondria evolved from a merger of an eubacterium with the eukaryotic ancestor.

The endosymbiont theory explains the origin of the organelles as an early intracellular (=endo) symbiotic relationship between eukaryotes and prokaryotes.

Nuclear genome sizes in plant species



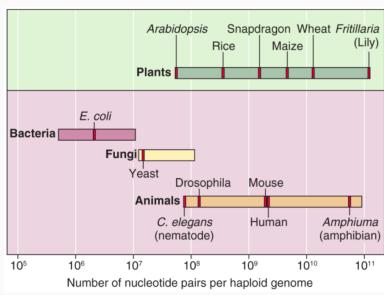
12 / 51

The condensed chromosomes of the five most important crop species and the model plant *Arabidopsis thaliana* shows large differences in genome size, which spans two orders of magnitude.

The very large genome of wheat is explained by polyploidization, which describes the fusion of genomes of different species.

The hexaploid genome of wheat is 3x the size of the diploid barley genome.

Comparison of genome sizes



[Jones et al., 2012]

Database of genome sizes:
<http://data.kew.org/cvalues>

13 / 51

The size of a genome is usually measured as the size of the **haploid genome**. This number is also called the **C-value**.

There is little relationship between the genome size and the taxonomic relationship of species.

Within each group of organisms the genome size varies by several orders of magnitude.

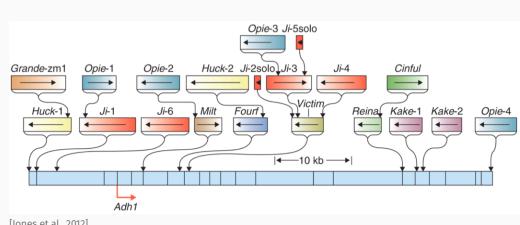
Within plants and animals, genome size variation is particularly large by more than 4 orders of magnitude.

What determines genome size?

In plants, most of the nuclear genome is made up of **repetitive DNA**.

Most repetitive DNA is made of **transposable Elements (TEs)**.

Diversity of transposable elements in the *Adh-1* region of maize:



14 / 51

The proportion of these repetitive elements differs strongly between plant genomes.

They consist of different types of genetic elements: tandem repeats, dispersed repeats, transposable elements (TEs).

The tandem repeats are short elements of few bp to tens of nucleotides that are repeated in large numbers as blocks.

Dispersed repeats are scattered throughout the genome.

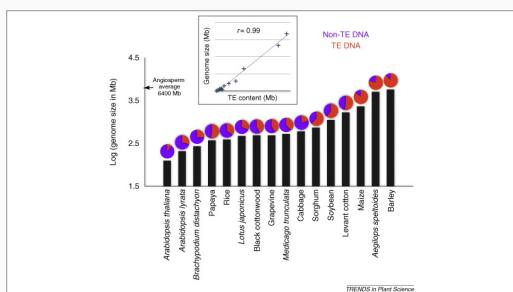
Transposable elements are 'jumping genes' that are able to move around in the genome and may cause new mutations.

All these genomic elements are characterized by a large diversity.

Because these genes are highly dynamic, they cause mutations and phenotypic variation and are likely important for plant breeding.

The figure shows the *Adh* gene region of maize that on a range of 60 kb contains only a single gene, but many other genetic elements.

Transposable elements (TE) and genome size



[Tenaillon et al., 2010]

15 / 51

The proportion of repetitive DNA in the nucleus was estimated for several plant species.

A comparison with the TE content and genome size shows a strong, positive correlation.

This observation suggests that genome size is mainly determined by the repetitive elements and that the non-repetitive proportion of genomes (which harbors the genes) can be quite similar between species of different genome sizes.

The eukaryotic cell

The genome of plants

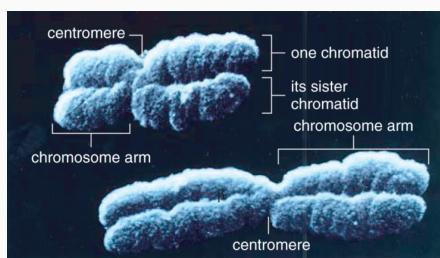
Organization of plant genomes

Replication of DNA

The structure and expression of genes

16 / 51

The eukaryotic nuclear genome is organized into chromosomes



[Jones et al., 2012]

Each chromatid ends in a **telomere**.

17 / 51

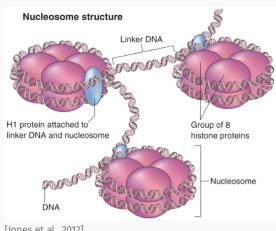
Centromeres are large and complex structures consisting of very long stretches of relatively short repeated sequences.

Telomeres are repeats of short sequences of few basepairs, in most plants of TTAGGG.

Chromosomes have numerous other distinct structural features, some of them with well distinct functions.

The organisation of chromosomes

- All eukaryotic chromosomes consist of DNA that are associated with proteins: **Chromatin**
- A large proportion of the proteins are **histones**
- The chromatin resembles the 'bead on a string' structure.



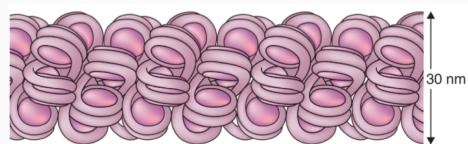
18 / 51

The DNA is packed up to several levels on the chromosome. The first level is the chromatin.

They are packed into octameric histone cores.

The organisation of chromosomes

Higher order coiling into a **solenoid** structure.



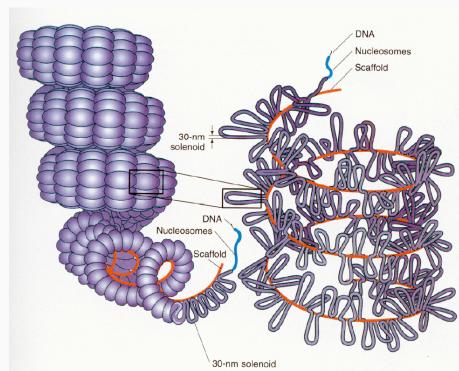
Solenoids + nuclear scaffold proteins → **chromatin loops**

19 / 51

Higher order coiling into a solenoid structure, which are 30 nanometer fibers, which are stabilised with Histone-1 proteins.

Solenoids can be further condensed by attaching to nuclear scaffold proteins to form chromatin loops

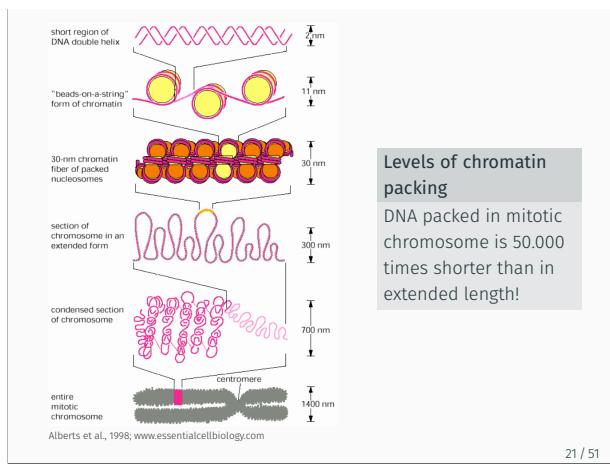
The organisation of chromosomes



20 / 51

This slide shows the formation of the nucleosomes as the next level of chromosomal organisation.

The higher order packing is achieved with additional types of scaffold proteins.

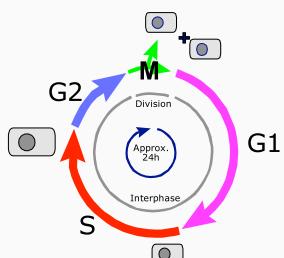


This figure shows the different level of integration and condensation that are required to condense the DNA molecule into a chromosome. It also shows the scale by which the DNA thread is compacted.

21 / 51

Division of a eukaryotic cell: Cell cycle

The growth of plants and their reproduction requires the division of cells



22 / 51

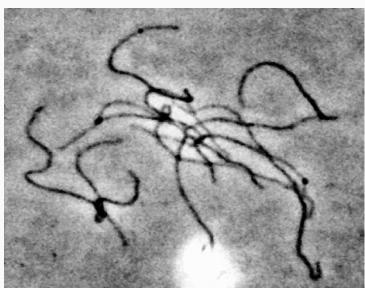
The growth of plants and their reproduction requires the division of cells.

The mitotic cell cycle consists of four stages that are characterized by different levels of chromosome condensation.

- Phase G1: Cell size increases.
- S (Synthesis) phase: DNA replication begins.
- Phase G2: Cell size increases again.
- M (Mitosis) phase: Cells stop growing and divide into two daughter cells. Mitosis is subdivided into prophase, prometaphase, metaphase, anaphase and telophase.

Division of cell chromosomes

Zea mays (L.) - Pachytene Chromosomes $2n = 20$



23 / 51

Pachytene chromosomes represent a certain stage in the cell division. It is the third stage in the prophase of the mitosis stage in which the chromosomes become shorter and thicker.

Division of cell chromosomes

Zea mays (L.) - Mitotic Chromosomes $2n = 20$



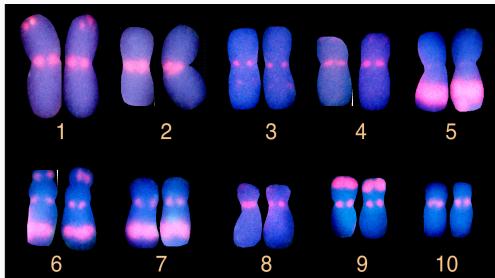
24 / 51

This figure shows mitotic chromosomes that are highly condensed. This is required because the chromosomes are distributed on the different cells.

Condensation is completed in the prometaphase.

Division of cell chromosomes

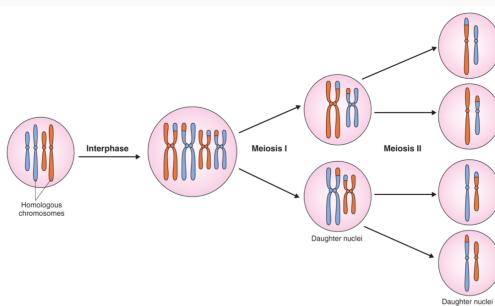
Zea mays (L.) - Mitotic Fluorescence labeled centromeres and knobs



25 / 51

In the highly condensed mitotic chromosomes, the different structural elements of chromosomes can be made visible with the appropriate techniques.

Meiosis produces haploid cells



[Jones et al., 2012]

26 / 51

Meiosis occurs in the reproductive tissues.

The role of meiosis is to produce haploid (1 chromosome set) gametes (sex cells, sperm or eggs) from diploid (2 chromosome sets).

It differs in some key aspects from mitosis:

- There are two cell divisions rather than one, and the four new resulting cells are haploid rather than diploid.
- Sister chromatids are held together in the prophase of meiosis I and undergo crossing over, which brings by recombination.

The eucaryotic cell

The genome of plants

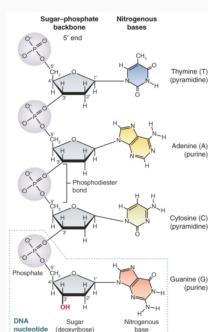
Organization of plant genomes

Replication of DNA

The structure and expression of genes

27 / 51

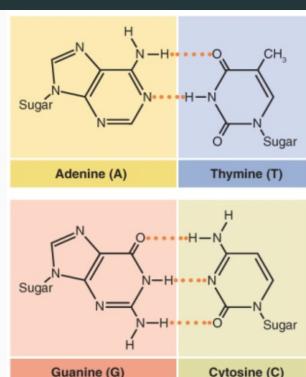
Building blocks of DNA



[Grotewold et al., 2015]

- DNA = Deoxyribonucleic acid

Building blocks of DNA



[Grotewold et al., 2015]

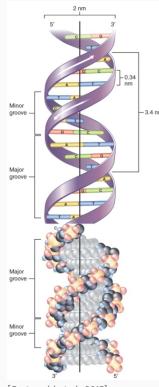
28 / 51

the building blocks of DNA (nucleotide) consist of three chemical compounds:

- The phosphate backbone, which links the sugars
- A deoxyribose sugar, which has a -H group at the 2' position (instead of an -OH group) in a ribose sugar
- A nitrogenous base. There are four different bases, of which two are purines and two are pyrimidines.

One pair of pyrimidine and purine base form hydrogen bonds.
The Adenine and Thymine form 2 hydrogen bonds.
The Guanine and Cytosine form 3 hydrogen bonds.

DNA is a double helix



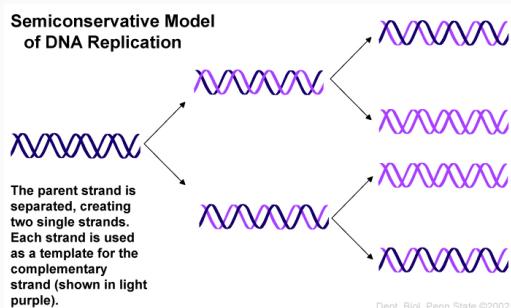
- DNA = Deoxyribonucleic acid
- Adenosine pairs with thymidine
- Cytosine pairs with guanine
- Long chain of individual pairs of nucleotides

[Grotewold et al., 2015]

30 / 51

The individual pairs of nucleotides can form very long chains. The three dimensional structure of the chain is of a double helix, which has a minor groove and a major groove.

Semiconservative Replication of DNA



31 / 51

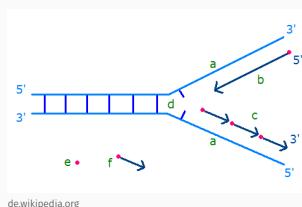
In semiconservative replication of DNA, the double helix is unwound into the two individual strands.

Then the complementary bases are synthesized on each strand to produce a complete double helix.

In the next replication step, again the DNA is unwound.

Because in each replication cycle, the template for the synthesis of the second strand is conserved from the previous cycle, the model is called semiconservative.

DNA replication at the replication fork



de.wikipedia.org

- a: Parent strands
- b: Leading strand
- c: Lagging strand
- d: Replication fork
- e: RNA-Primer
- f: Okazaki-Fragment

32 / 51

The directed nature of the DNA strands allow synthesis of the complementary strand only in one direction.

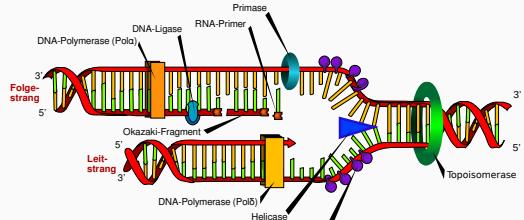
For this reason the mechanisms of DNA replication differs between the two directions at the replication fork.

In the leading strand, DNA synthesis can move on with the replication fork from right to left direction.

In the opposite lagging strand, small RNA primers have to anneal to the strand and prime the synthesis of short fragments, which are called Okazaki fragments after their discoverer.

A complex molecular machinery for DNA replication

Many different proteins are required for DNA replication



33 / 51

The replication of DNA involves a complex molecular machinery of numerous proteins.

An enzyme called topoisomerase or helicase unwinds the DNA.

Single strand binding proteins (SSB) bind to the single strands and keep them apart from each other.

The two strands are called leading and lagging strand.

Small fragments of DNA bind as primers to the open single strand. In the lagging strand, primers called 'Okazaki fragment' have to prime repeatedly.

A **DNA polymerase** attaches to the DNA and synthesizes the second, complementary strand on each single strand.

A **DNA ligase** links the pieces of newly synthesized fragments on the lagging strand.

The eucaryotic cell

The genome of plants

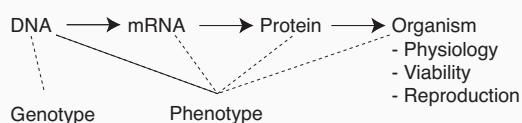
Organization of plant genomes

Replication of DNA

The structure and expression of genes

34 / 51

The Central Dogma of Molecular Biology



35 / 51

The Central Dogma of Molecular Biology states that information flows only in one direction.

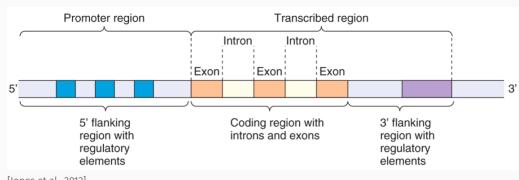
The information encoded in DNA is transcribed into RNA and then translated into proteins.

No information from proteins is brought back to RNA or to DNA.

One exception is a flow back from RNA to DNA via epigenetic processes.

An important consequence of the dogma is that genetic variation in DNA causes variation in proteins, but variation in proteins does not lead to variation in DNA.

Structure of plant genes



[Jones et al., 2012]

36 / 51

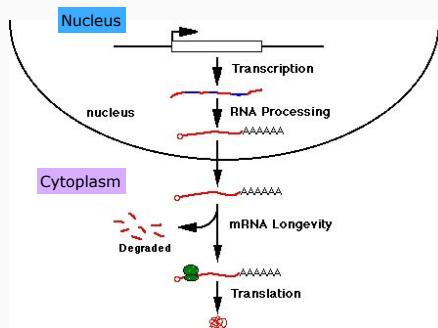
Typical plant genes have a mosaic structure.

The first level is the promotor region (or regulatory region) and the transcribed region.

The transcribed region is transcribed into ribonucleic acid (RNA), which is processed and translated into protein.

The transcribed region consists of a sequence of exons and introns, and of a 'tail' of a flanking region with regulatory elements.

Eukaryotic genes: Transcription



37 / 51

The conversion of information encoded in genes into a messaging molecule (mRNA) is called transcription.

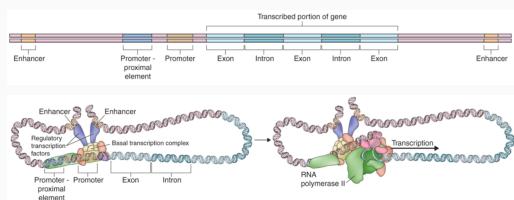
A RNA molecule is transcribed from the DNA containing information from introns to exons.

The RNA molecule is processed by removing the noncoding introns and adding a polyA tail and a cap to protect the molecule from degradation.

The resulting messenger RNA (mRNA) is transported from the nucleus into the cytoplasm.

Depending on cellular signals, the RNA is either degraded by enzymes or translated into a protein.

Initiation of transcription



[Jones et al., 2012]

38 / 51

The initiation of transcription is a complex process that involves several components.

Promotor and other regulatory elements control the timing and the level of gene transcription.

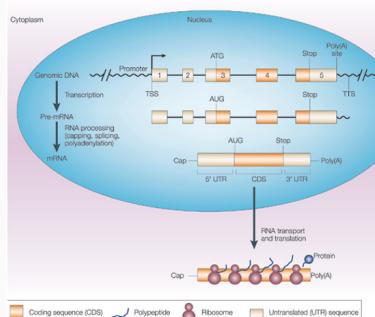
There are several proteins that bind to the different types of regulatory elements. These proteins are called **transcription factors**. They can initiate or repress the expression of a certain gene. They bind to regulatory sequences such as enhancers and promotor elements and form protein complexes with the transcription proteins.

Some transcription regulators bind to response elements in reaction to environmental conditions such as nutrient status, environmental stress, etc.

A group of other proteins, which bind to the so-called core promotor elements. They include the TATA-box binding proteins and RNA polymerases that synthesize the RNA molecule.

The combination of regulatory proteins and polymerases and accessory proteins is called the transcription initiation complex.

Eukaryotic genes: RNA Processing



39 / 51

RNA processing is an important step in transcription.

The newly synthesized RNA molecule consisting of exons, introns and the tail of the flanking region, are modified.

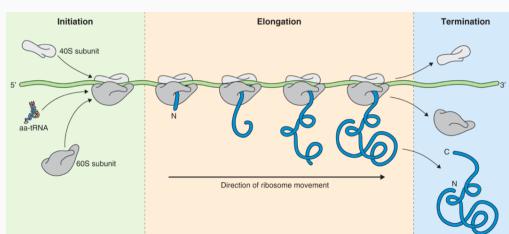
The resulting molecule is called messenger RNA or mRNA.

A poly-A tail is added to the 3' end of the mRNA precursor to protect it from degradation.

A methyl-Guanosin 5' cap is also added for protection.

By a process called splicing, the introns are removed to produce the mature mRNA.

Protein synthesis: Translation of mRNA into proteins



40 / 51

Protein coding genes are translated into proteins from their mRNA intermediate.

This is achieved by a complex of ribosomes, mRNA and transferRNA molecules.

Ribosomes are protein complexes, which bring mRNA and tRNA molecules together. They consist of two subunits the 40S and 60S subunits.

tRNA molecules are molecules to which an amino acid is coupled. The proteins are synthesized by removing the amino acid from the tRNA and adding them to the nascent protein by proteins called peptyl transferases.

The genetic code of protein synthesis

Amino acid	3-Letter code	1-Letter code	Codons
Alanine	Ala	A	GCC, GCU, GCG, GCA
Arginine	Arg	R	CGC, CGG, CGU, CGA, AGA, AGG
Asparagine	Asn	N	AAU, AAC
Aspartic acid	Asp	D	GAU, GAC
Cysteine	Cys	C	UGU, UGC
Glutamic acid	Glu	E	GAAG, GAG
Glutamine	Gln	Q	CAA, CAG
Glycine	Gly	G	GGU, GGC, GGA, GGG
Histidine	His	H	CAU, CAC
Isoleucine	Ile	I	AUU, AUC, AUA
Leucine	Leu	L	UUA, UUG, CUA, CUG, CUU, CUC
Lysine	Lys	K	AAA, AAG
Methionine	Met	M	AUG
Phenylalanine	Phe	F	UUU, UUU
Proline	Pro	P	CCU, CCC, CCA, CCG
Serine	Ser	S	UCU, UCC, UCA, UCG, AGU, AGC
Threonine	Thr	T	ACU, ACC, ACA, ACG
Tyrosine	Tyr	Y	UAU, UAC
Tryptophan	Trp	W	UGG
Valine	Val	V	GUU, GUC, GUA, GUG
"Stop"	—	—	UAA, UAG, UGA

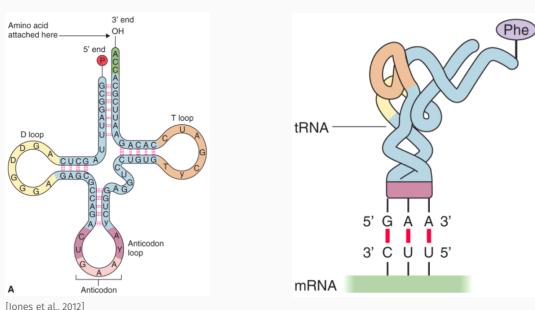
[Jones et al., 2012]

41 / 51

The genetic code relates the codons, which are triplets of nucleotides in the mRNA to the amino acids in the protein.

The genetic code is degenerate, which means that some amino acids are encoded by more than one triplets of nucleotides (or codons). For example leucine is encoded by 6 codons.

Structure and pairing of transfer RNAs (tRNAs)



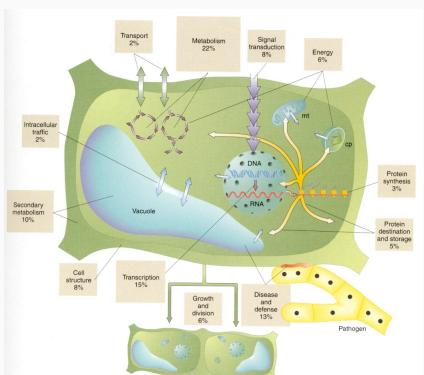
[Jones et al., 2012]

42 / 51

tRNAs have a distinct cloverleaf structure that consists of three loops. Based on these loops, it undergoes a distinct folding in three dimensions.

The three nucleotides at the tip of the anticodon loop recognize the complementary codon on the mRNA, which helps to align and pair the tRNA and the amino acid coupled to it to the peptidyl transferase that synthesizes the protein.

Functions of genes

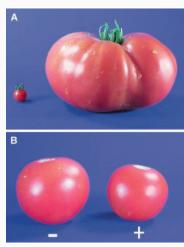


43 / 51

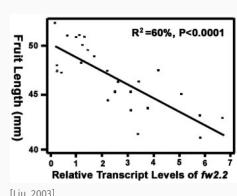
The average plant genome has between 20,000 to 40,000 protein-coding genes. They have very different functions for the plant. The largest proportion of genes encodes proteins, which play a role in metabolism (for example, enzymes).

The second largest groups of genes are involved in preventing disease and defending the plant from pests and pathogens.

The fw2.2 gene controls fruit size in tomato



[Frary et al., 2000]



[Liu, 2003]

44 / 51

The fw2.2 gene was identified in a QTL study. A genetic variant from a wild tomato that was introduced into a modern tomato variety reduced the fruit size by 30%.

Subsequent analyses showed that the cause for the differences is not the protein sequence, but changes in the regulatory regions that control the expression level of the gene.

Changes in the expression level of the gene showed that there is a strong correlation with fruit size and the amount of mRNA produced by this gene.

The function of the gene is the control of cell division. Fruits with the wild fw2.2 variant have fewer cell divisions and therefore smaller fruits, which explains the negative correlation between transcript level and fruit length.

This work shows that there is a direct relationship between a regulatory mutation, gene expression and yield.

Summary

- Plants consists of cells, which contain a diversity of organelles
- Plants have nuclear and organellar genomes
- Cell division and genome replication are fundamental cellular processes
- Genetic variation of genes controlling these processes may affect agronomic traits

45 / 51

Further reading

- Any introductory genetics textbook
- Jones et al.: *The Molecular Life of Plants* (2012), Wiley
- Grötewold, Chappel, Kellogg: *Plant Genes, Genomes and Genetics* (2015), Wiley
- Griffiths, Wessler, Carroll, Doebley: *Introduction to Genetic Analysis* 10th edition, Freeman

46 / 51

Review questions i

1. Why is a basic understanding of cell and molecular biology important for plant breeding?
2. What are the key differences between a prokaryotic and an eukaryotic cell?
3. How is the genetic information encoded in the genome read out and used?
4. Why do chloroplasts and mitochondria have their own genomes? What are the most important differences to nuclear genomes?
5. What are possible reasons why the size of plant genomes is so variable between species?

47 / 51

Review questions ii

6. What are main causes of differences in genome size and numbers of genes between the different organismal groups?
7. Which functional groups of genes are more frequent than others?
8. Which approaches are used by the cell to condense the long DNA molecule into a highly compact chromosome?
9. What is the advantage of semiconservative replication in cell division?
10. What are key characteristics of transcription and translation?
11. What is the difference between transfer RNA and messenger RNA?

48 / 51

Review questions iii

12. Which different types of regulatory elements exist, and what are their key functions?
13. How can basal cellular processes affect complex traits such as crop yield?

49 / 51

References i

- [img] Frary, A., Nesbitt, T. C., Frary, A., Grandillo, S., Van Der Knaap, E., Cong, B., Liu, J., Meller, J., Elber, R., Alpert, K. B., and others (2000). fw2.2: a quantitative trait locus key to the evolution of tomato fruit size. *Science*, 289(5476):85–88.
- [img] Griffiths, Antony J. F., Wessler, Susan R., Carroll, Sean B., and Doebley, John (2012). *Introduction to Genetic Analysis*. W. H. Freeman and Company, 10th edition.
- [img] Grötewold, E., Chappell, J., and Kellogg, E. A. (2015). *Plant genes, genomes and genetics*. John Wiley & Sons.
- [img] Jones, R., Ougham, H., Thomas, H., Waaland, S., et al. (2012). *Molecular life of plants*. Wiley-Blackwell.
- [img] Kawashima, T. and Berger, F. (2014). Epigenetic reprogramming in plant sexual reproduction. *Nature Reviews Genetics*, 15(9):613–624.

50 / 51

References ii

-  Liu, J. (2003). Generation and Analysis of an Artificial Gene Dosage Series in Tomato to Study the Mechanisms by Which the Cloned Quantitative Trait Locus fw2.2 Controls Fruit Size. *PLANT PHYSIOLOGY*, 132(1):292–299.
-  Tenaillon, M. I., Hollister, J. D., and Gaut, B. S. (2010). A triptych of the evolution of plant transposable elements. *Trends in Plant Science*, 15(8):471–478.

51 / 51



Plant Biotechnology and Genetic Engineering

3501-211 Genetische Grundlagen der Pflanzenzüchtung

Prof. Karl Schmid
WS 2023/2024

Institute of Plant Breeding, Seed Science and Population Genetics
University of Hohenheim

1 / 45

Overview

Background

Tissue culture

Genetic transformation and GMOs

Genome editing: A new breeding technology

2 / 45

Background

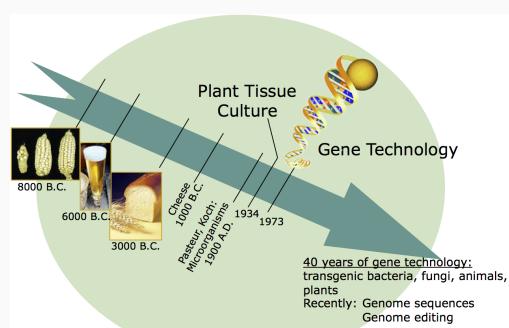
Tissue culture

Genetic transformation and GMOs

Genome editing: A new breeding technology

3 / 45

10,000 years of biotechnology



4 / 45

Biotechnology in the broad sense has been used by humans since almost 10,000 years and falls together with the beginning of agriculture, and plant and animal domestication.

In that respect biotechnology can also be described as the domestication of microbes and yeasts that were used in the beginning of agriculture to make beer, wine, cheese and other products from milk.

Definition of biotechnology

In a broad sense:

Plant biotechnology covers many of the tools and techniques that are commonly used in agriculture and food production.

In a narrow sense:

Biotechnology considers only the new DNA techniques, molecular biology and reproductive technological applications, like gene manipulation, gene transfer, DNA genotyping and cloning.

5 / 45

Biotechnology in the broad sense has been used for thousands of years. It did not require an understanding of the underlying molecular and genetic processes.

Modern Biotechnology is based on a deep mechanistic and causal understanding of the processes that occur in the cell and in organisms. They are manipulated using a directed (or rational) approach in order to achieve the desired result or product.

This is a fundamental difference to the old, or classical biotechnology.

Definition of biotechnology

Biotechnology defines the use of biological organisms or processes in any technological application

Classical biotechnology

Production of wine, beer, cheese, bread

Modern biotechnology

E.g., cell fusion, organisms with recombinant DNA, functional food, pharmaceuticals

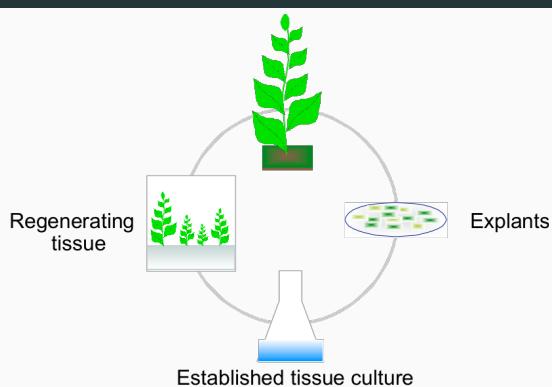
6 / 45

The methods of modern biotechnology are fundamentally different from the classical biotechnology.

In modern biotechnology, methods that in particular influence the reproduction of organisms and methods of molecular biology are being used to modify living organisms.

In classical biotechnology, these methods relied mainly on the selection of organisms (e.g., strains of yeast) which differ in their phenotypic characteristics (e.g., produce a better cheese).

Omnipotency of plant cells



7 / 45

Plant biotechnology strongly depends on the omnipotency of plant cells, because for many biotechnological methods, plants need to be regenerated from small pieces of plant tissues (e.g., leaves), which are transformed and manipulated.

The omnipotency of plants can be induced by applying certain nutrients and plant hormones, because many explants are created from specialised tissues.

Therefore, the omnipotency is created from cells, which already have acquired a particular fate or state (e.g., as a leaf cell).

Background

Tissue culture

Genetic transformation and GMOs

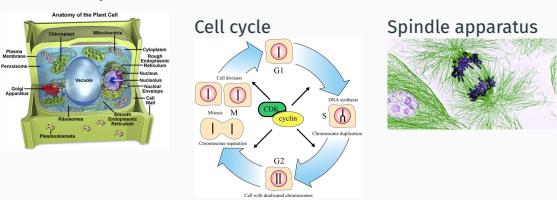
Genome editing: A new breeding technology

8 / 45

Tissue culture : Key technology for plant biotechnology

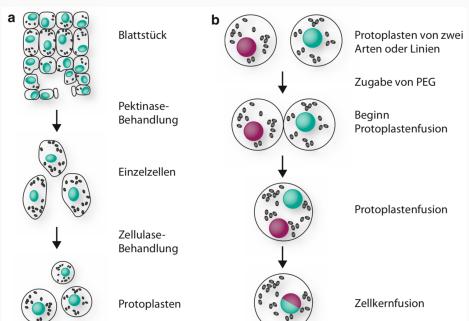
Tissue culture requires an understanding and manipulation of basic cellular processes.

Cell anatomy



9 / 45

Protoplast culture



[Kempken, 2020]

10 / 45

Protoplast fusion

Merger of two cells, whose cell walls were dissolved with enzymes



http://de.wikipedia.org/wiki/Datei:Protoplast_fusion.jpg

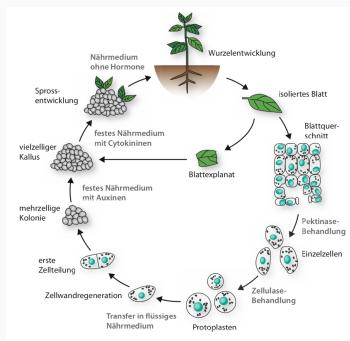
After cell fusion, the fusion of nuclei frequently occurs
(**karyogamy**)

Results are **somatic hybrid** plants, which are **polyploid**

Use in breeding: Eg. potato is more resistant against viruses

11 / 45

Regeneration from protoplasts



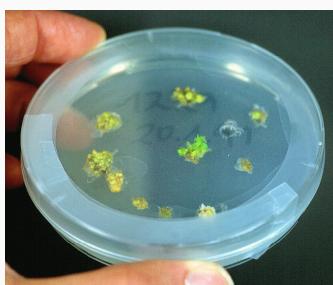
[Kempken, 2020]

12 / 45

Shoot regeneration on solidified medium

Cells need to be regenerated into whole plants

Eg., hop (*Humulus lupulus*):



Gerd Weber, University of Hohenheim

13 / 45

Basic methods of tissue culture

Tissue culture is the cultivation of plant cells or the regeneration of whole plants from plant cells

- Starting of plant tissue cultures
- Tissue culture laboratory
- Culture media

14 / 45



Plant tissue culture laboratory

Laminar flow hood - A sterile work environment:

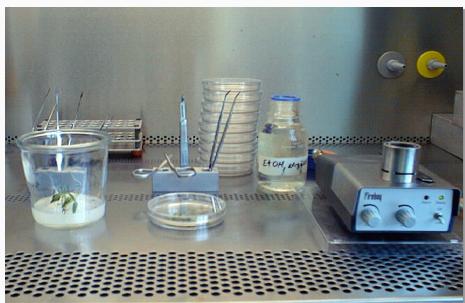


Photo: Gerd Weber, University of Hohenheim

15 / 45

- Laminar flow hood
- Composition of plant tissue culture media
- Preparation of media; sterilization
- Tools and culture vessels
- Growth chambers
- Observation and documentation

Background

Tissue culture

Genetic transformation and GMOs

Genome editing: A new breeding technology

16 / 45



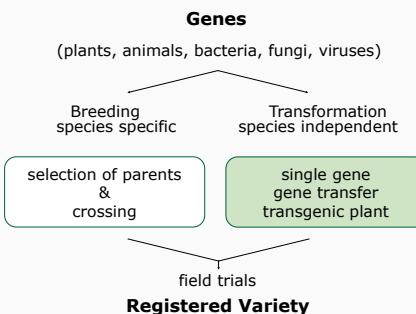
Genetically Modified Organisms - GMO

Definition of a GMO:

The genetic material of an organism has been modified in a way not possible by crossing or recombination

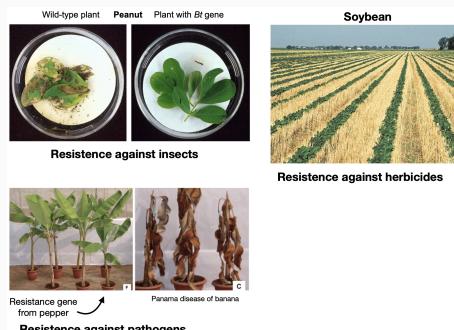
17 / 45

Difference between breeding and gene transfer



18 / 45

Examples of genetically engineered, transgenic plants



20 / 45

Methods for gene transfer

Vector-mediated gene transfer *Agrobacterium tumefaciens* - Ti Plasmid

Direct gene transfer

- Particle gun
- DNA uptake by protoplasts

21 / 45

Natural infection with Agrobacterium



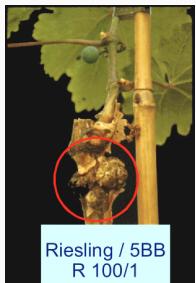
Foto: Gerd Weber, University of Hohenheim

22 / 45

Agrobacterium tumefaciens

Targeted production of transgenic tissues / plants by Agrobacterium infection of a wine rootstock.

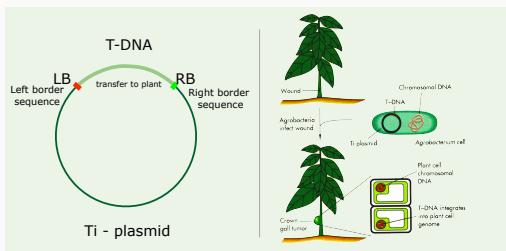
Transgenic Tissue



Gerd Weber, University of Hohenheim

23 / 45

Agrobacterium tumefaciens - Gene transfer with Ti plasmid

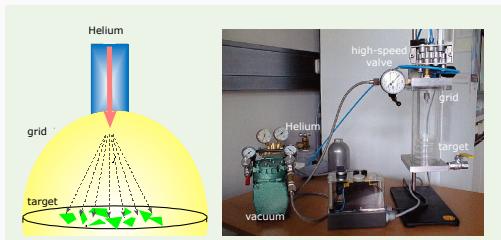


Gerd Weber, University of Hohenheim

24 / 45

In the case of Agrobacterium infected tissue the bacterium is the vector of the Ti plasmid, which is then brought into plant cells. The plasmid is then integrated into the plant genome in random positions using the enzymes encoded by the Ti plasmid, which are produced in the plant cell using the "machinery" of the host cell.

Direct gene transfer - Particle Gun



Gerd Weber, University of Hohenheim

25 / 45

Not all plants can be transformed with Agrobacterium because they are resistant against it.

For these species other methods need to be used such as a particle gun.

Very small particles (e.g. gold or ceramics) are covered with the Ti plasmid or other DNA and shot onto plant tissue. The particles enter into the cell and the DNA is subsequently released and the enzymes for integration are being produced.

Particle gun: GUS expression in maize embryos

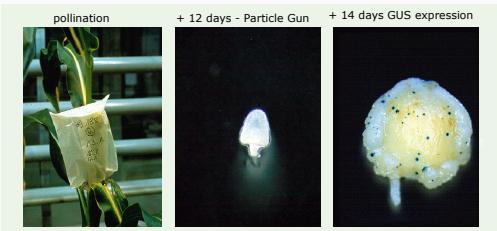


Foto: Gerd Weber, University of Hohenheim

26 / 45

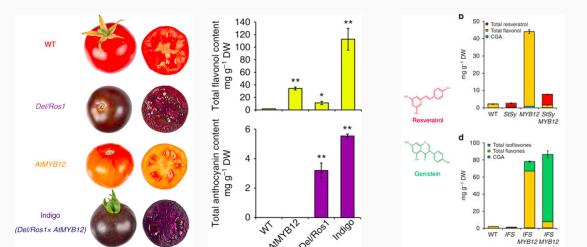
GUS is a reporter gene system used to monitor the expression of transgenes.

If the transgenes are not integrated into the host genomes, expression is said to be transient, because the transgene is not inherited into the next generation.

If the transgenes integrate, expression is stable and inherited into the next generation.

Example: Genetic engineering of tomato

Goal: Produce tomato with a high concentration of secondary metabolites [Zhang et al., 2015]



27 / 45

The goal of the experiment to produce new tomato varieties with favorable properties that increase the nutritional value.

TECH TIMES PERSONAL TECH BIZ TECH FUTURE TECH SCIENCE LIFE T-LOUNGE

TAG AtMYB12, Super Tomato, Genetically Modified Fruit, Tomato Juice

Genetically Modified Super Tomato May Help Fight Cancer, Diabetes And Alzheimer's Disease

By Katherine Dera, Tech Times | October 28, 3:29 AM

Like **Follow** **Share(2276)** **Tweet(40)** **Reddit** **Comments(0)** **Subscribe**



A study led by John Innes Centre researchers in the United Kingdom (UK) resulted in the mass production of natural compounds in a single, genetically modified fruit - a tomato. These compounds include life-extending Resveratrol found in wine and Genistein found in tofu which has cancer-preventing benefits.

By introducing a protein called AtMYB12, which is normally found in a garden weed called thale cress (*Arabidopsis thaliana*), scientists found that the protein activates a wide set of genes responsible for natural compound production in the tomato plant. The AtMYB12 acts like a plug or tap that scientists can control to

28 / 45

Since genetic engineering is viewed very critically by society, the GMO plants are often marketed as a great scientific progress. However, such a marketing as "superfood" is not without risks because it often leads to claims and exaggerations that are not true or difficult to proof.

Natural GMO plants

- Ti-Plasmid-DNA like genes were found in 39 of 275 dicotyledonous species (14%)
- Also in monocots: greater yam and banana

Plant Molecular Biology
https://doi.org/10.1007/s11103-019-00913-y

Widespread occurrence of natural genetic transformation of plants by Agrobacterium

Tatiana V. Matveeva¹ · Léon Otten² 

Received: 18 June 2019 / Accepted: 21 August 2019
© Springer Nature B.V. 2019

Abstract
Key message Naturally transgenic plant species occur on an unexpectedly large scale.

29 / 45

In the discussion of genetic engineering, one frequent argument by critics is that it is 'unnatural'. However, more and more evidence shows that this is not true anymore. Naturally transgenic GMO plants that were infected by Agrobacterium and have inserted bacterial genes, which are expressed in the host plant are quite frequent. A recent study found that about 10% of 234 investigated plant species are natural transgenics!

Sweet potato is a natural GMO

PNAS The genome of cultivated sweet potato contains *Agrobacterium* T-DNAs with expressed genes: An example of a naturally transgenic food crop

Tina Kyndt¹, Dore Quispel^{1,2}, Hong Zhai³, Robert Janss⁴, Marc Ghislain⁵, Qingshang Liu⁶, Godfrey Gheysen⁷, and Jan F. Krens¹

¹ Institute of Plant Biotechnology, Ghent University, 9000 Gent, Belgium; ² International Potato Center, Lima 02, Peru; ³ Beijing Key Laboratory of Crop Genetic Improvement, Laboratory of Crop Genetics and Breeding, Ministry of Education, China Agricultural University, Beijing, China; ⁴ Laboratory of Crop Genetic Resource CnC, Department of Agriculture, Research Institute of Ghent University, Ghent, Belgium; ⁵ Department of Plant Biotechnology, Ghent University, 9000 Gent, Belgium; ⁶ Department of Plant Pathology, China Agricultural University, Beijing, China; ⁷ Department of Plant Pathology, Ghent University, 9000 Gent, Belgium



30 / 45

Background

Tissue culture

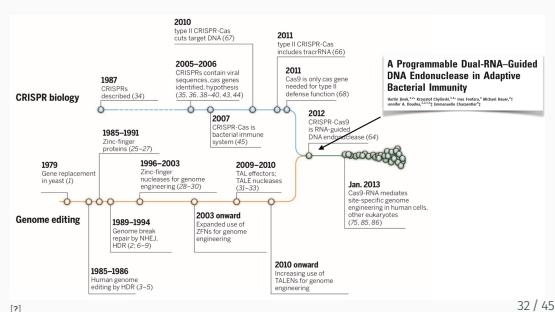
Genetic transformation and GMOs

Genome editing: A new breeding technology

31 / 45

Timeline: Development of CRISPR/Cas9 technology

The investigation of the began in studies of bacterial immunity against viruses that infect bacteria (phages).



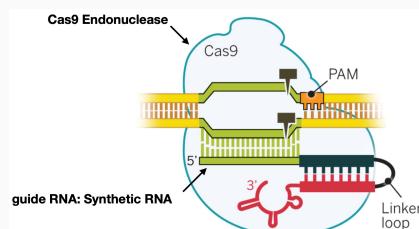
[?]

32 / 45

Genome editing is precision mutagenesis

CRISPR/Cas9 is a **programmable** DNA scissor

- **CRISPR:** Clustered Regularly Interspaced Short Palindromic Repeats
- **Cas:** CRISPR-associated protein



Modified after [?]

33 / 45

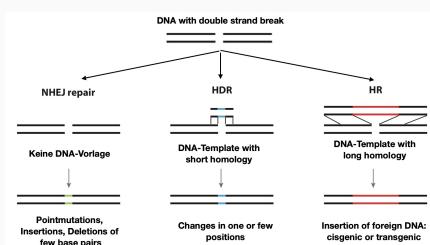


THE ROYAL SWEDISH ACADEMY OF SCIENCES

[?]

34 / 45

Consequences of double strand break by Cas9 protein



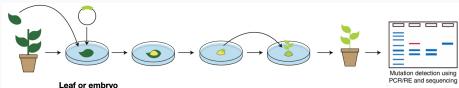
- NHE: Non-homologous end-joining
- HDR: Homology-directed repair
- HR: Homologous recombination

Genome editing allows to create novel genetic variation in predefined genomic regions

35 / 45

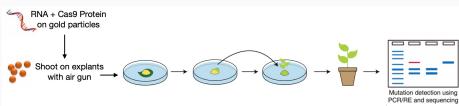
Different variants of genome editing

Transgenic: T-DNA vector with genes for Cas9 and guide RNA, Transformation with *Agrobacterium*



T-DNA encodes guide RNA and Cas-9 protein

Using particle bombardment:



36 / 45

mRNA encodes guide RNA and Cas-9 protein

Comparison of GMO and Non-GMO based genome editing

Target site	Target site sequence with PAM*	Cas9 only (%)	DNA delivery (%) [†]	RNP delivery (%) [†]
MS45 MS45 off-site	GGCCGAGGTGACTACCGGCCG GCCGGAGGTGACTACCGGCCG	0.002 0.002	0.34 0.18	0.69 0.01

Non transgenic (Ribonucleic acid coated gold particles)

GCTGCCCGAGGTGACTACCGGCCG WT
GCTGCCCGAGGTGACTACCAAGGCCGG +1
GCTGCCCGAGGTGACTACCTGGCCGG +1
GCTGCCCGAGGTGACTACCCGGCCGG +1
GCTGCCCGAGGTGACTACCGGCCGG -1



Transgenic (DNA)

GCTGCCCGAGGTGACTACCGGCCG WT
GCTGCCCGAGGTGACTACCAAGGCCGG +1
GCTGCCCGAGGTGACTACCTGGCCGG -1
GCTGCCCGAGGTGACTACCCGGCCGG +1
GCTGCCCGAGGTGACTACCGGCCGG +1

[?]

Both methods produce same outcome, but non-GMO editing has 2x higher efficiency.

37 / 45

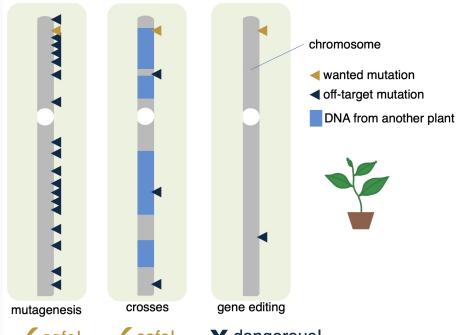
Initial rapid development of genome edited plants

Pflanze	Merkmal	Technik	Anwendungsbezug und Funktionsnachweis	bereits von Regulierung angenommen in
verbesserte Nahrungs- bzw. Futtermittelequalität				
Luzerne	verringelter Ligningehalt	TALEN	APHIS-Datenbank	USA
Kartoffel	verringerte Acrylamidbildung	TALEN	Claßen et al. 2016	USA
Leinodter	verbesserte Fettsäurezusammensetzung	CRISPR-Cas	Ozeyhan et al. 2018	Chile
Salat	erhöhte Vitamin-C Gehalte	CRISPR-Cas	Zhang et al. 2018	-
Soja	verbesserte Fettsäurezusammensetzung	TALEN	APHIS-Datenbank	USA, Chile (Vermarktung)
Weizen	verbesserter Faseranteil	TALEN	APHIS-Datenbank	USA
geringerer Glutengehalt				
Weizen	CRISPR-Cas	Sánchez-Léón et al. 2018	-	
Reduktion des Pestizideinsatzes, Wasserverbrauchs und von Ernteverlusten				
Mais	Pilzresistenz	CRISPR-Cas	APHIS-Datenbank	USA
	Trockenobohrhardt	CRISPR-Cas	Shi et al. 2017	-
Kakao	Pilzresistenz	CRISPR-Cas	Fister et al. 2018	-
Soja	Trockenobohrhardt	CRISPR-Cas	APHIS-Datenbank	USA
Tomate	Bakterienresistenz	CRISPR-Cas	Thomasset et al. 2016	-
Reis	Wurzelkrebsresistenz	CRISPR-Cas	APHIS-Datenbank	USA
	Salztoleranz	CRISPR-Cas	Duan et al. 2016	-
Weizen	Pilzresistenz	TALEN	Wang et al. 2014	USA
Banane	Pilzresistenz	CRISPR-Cas	Dale et al. 2017	-
Maniok	Virusresistenz	CRISPR-Cas	Gómez et al. 2019	-

Source: <http://www.leopoldina.org>

38 / 45

Differences in regulation between breeding methods



Modified after Etienne Bucher, Agroscope, Switzerland

39 / 45

Summary i

- Plant biotechnology comprises tissue culture, the manipulation of plant tissues, and genetic engineering
- The development and application of biotechnological methods such as protoplast fusion requires a thorough understanding of cell anatomy, the cell cycle and cell physiology.
- Many different plant traits can be changed by biotechnology that include resistances, food quality, pharmaceuticals and renewable resources
- The two most important methods for gene transfer are vector-mediated gene transfer via *Agrobacterium tumefaciens* or direct gene transfer via particle gun or DNA uptake by protoplasts

40 / 45

Summary ii

- Genetic transformation produces products with a high concentration of favorable metabolites
- There are many natural GMOs that originated by infection with *A. tumefaciens* and subsequent transfer of DNA
- Genome editing is a new breeding method that is based on the targeted introduction of double strand breaks, whose repair induces new mutations

41 / 45

Further reading i

- Deutsche Forschungsgemeinschaft DFG: Grüne Gentechnik (https://www.dfg.de/dfg_magazin/aus_gremien_politikberatung/gruene_gentechnik/publikationen_gruene_gentechnik/index.html)

42 / 45

Review questions i

1. What are the key differences between genetic engineering and classical plant breeding?
2. Which type of traits are particularly suitable for genetic engineering?
3. Why is tissue culture defined as biotechnology?
4. Why are quantitative traits less suitable for genetic engineering?
5. Why are marker genes required in genetic engineering?
6. Describe one trait in greater detail that has been created or improved by genetic engineering: Which gene(s) were transformed and what was the source organism?
7. Explain the basic principle of genetic transformation with *Agrobacterium tumefaciens*.

43 / 45

Review questions ii

8. What is the function of the Ti Plasmid in *Agrobacterium tumefaciens*-mediated gene transformation?
9. What is a naturally transgenic plant and how does one arise?
10. Why can genome editing be called mutagenesis and genetic engineering at the same time?
11. After cutting DNA with CRISPR/Cas9, what natural mechanism plays an important role in generating new genetic variation?
12. What are the differences in the number and nature of mutations between cross-breeding, classical mutagenesis and genome editing?
13. How can the targeted introduction of mutations lead to the development of resistant plant varieties?

44 / 45

References i

-  Kempken, F. (2020). *Gentechnik bei Pflanzen - Chancen und Risiken*. Springer Spektrum, 5th edition.
-  Zhang, Y., Butelli, E., Alseekh, S., Tohge, T., Rallapalli, G., Luo, J., Kawar, P. G., Hill, L., Santino, A., Fernie, A. R., and Martin, C. (2015). Multi-level engineering facilitates the production of phenylpropanoid compounds in tomato. *Nature Communications*, 6(1):8635.

45 / 45



Mendelian Rules of Inheritance

3501-211 Genetische Grundlagen der Pflanzenzüchtung

Prof. Karl Schmid
WS 2023/2024

Institute of Plant Breeding, Seed Science and Population Genetics
University of Hohenheim

1 / 38

Overview

Background

Mendel's first rule: Equal segregation of variation

Mendel's second rule: Independent Assortment of Traits

2 / 38

Background

Mendel's first rule: Equal segregation of variation

Mendel's second rule: Independent Assortment of Traits

3 / 38

Recapitulation

- Basic structure of plant cells
- Genome and genes; genome size
- Transcription and translation
- Replication and Meiosis/Mitosis

Today: 'Mendelian' or 'classical genetics'

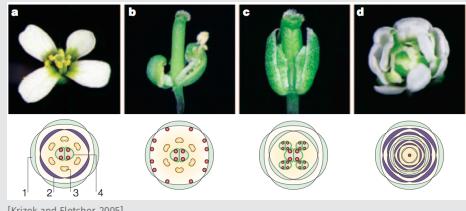
4 / 38

In the previous two lectures, we discussed the cellular and molecular aspects of plant life and their application in biotechnology. In this lecture, we will discuss what is called 'classical' or Mendelian genetics because it dates back to the work of Gregor Mendel. Our goal is to understand the principle of inheritance and the relationship between genes and traits.

Single gene inheritance

- Focus on traits that controlled by **single genes**
- Original trait is called **wild type**, the modified trait a **mutant**.

Arabidopsis thaliana flower mutants



5 / 38

Single gene inheritance is best investigated by analysis crosses of plants that differ in a trait controlled by a single gene and the analysis of inheritance patterns.

This is the experiment what Mendel did, although he did not know about the underlying genetic basis of the inheritance of his traits.

Background

Mendel's first rule: Equal Segregation of Variation

Mendel's second rule: Independent Assortment of Traits

6 / 38

Gregor Mendel (1822 - 1884)



- Catholic monk
- Amateur scientist
- 1856 - Beginning of pea experiments
- 1866 - Publication: *Versuche über die Pflanzenhybriden*
- Experiments also with other plant species

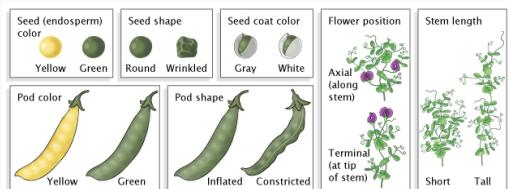
7 / 38

Mendel was the first one to use single-gene inheritance to discover a gene.

He chose the garden pea, *Pisum sativum* as his model species because it had many advantages and varieties that differed in several traits were available.

Traits investigated by Mendel

- Alternative word for **property**: **character** or trait
- A modern word for this is **phenotype**.



<http://www.nature.com/scitable/topicpage/gregor-mendel-and-the-principles-of-inheritance-593>

8 / 38

Mendel worked with mutants that were discovered by others, and he wanted primarily identify patterns of inheritance.
Mendel investigated seven properties of peas.

Pure lines

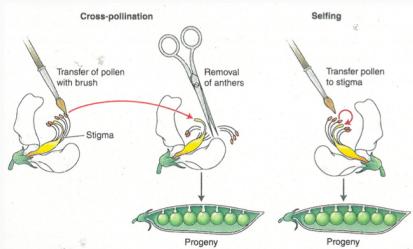
- Mendel worked with **pure lines**, which are **homozygous lines**.
- “Pure”: Crosses of individuals from the same line will produce homogeneous offspring.

9 / 38

Mendel worked with pure lines:
These are homozygous (reinerbige) lines.
Crosses of individuals from the same homozygous line will produce offspring that is homogeneous.

Types of crosses

- Cross-pollination of two individuals
- Self-fertilization or selfing

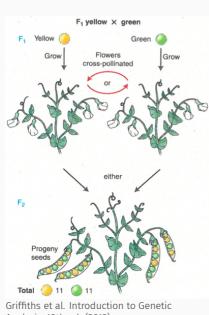


Griffiths et al. Introduction to Genetic Analysis, 10th ed. (2012)

10 / 38

Mendel used two types of crosses: Cross-pollination of two individuals that differed in a trait, or self-pollination (or selfing) of a single individual.

Cross-pollination



Griffiths et al. Introduction to Genetic Analysis, 10th ed. (2012)

- Parental generation (P): The two lines that are crossed
- Filial generation (F₁): the resulting offspring
- F₁ Generation can be crossed among each other, or self-fertilized
- Result: 1:1 phenotypic ratio in the F₂ generation

The figure show an example of a cross-pollination.
A surprising result of this mating scheme was that the green pea color was not present in the F₁, but re-appeared in the F₂ generation.

11 / 38

Self-pollination



Griffiths et al. Introduction to Genetic Analysis, 10th ed. (2012)

12 / 38

Results of Mendel's crosses

Table 2-1 Results of All Mendel's Crosses in Which Parents Differed in One Character

Parental phenotypes	F ₁	F ₂	F ₂ ratio
1. round × wrinkled seeds	All round	5474 round; 1850 wrinkled	2.96:1
2. yellow × green seeds	All yellow	6022 yellow; 2001 green	3.01:1
3. purple × white petals	All purple	705 purple; 224 white	3.15:1
4. inflated × pinched pods	All inflated	882 inflated; 299 pinched	2.95:1
5. green × yellow pods	All green	428 green; 152 yellow	2.82:1
6. axial × terminal flowers	All axial	651 axial; 207 terminal	3.14:1
7. long × short stems	All long	787 long; 277 short	2.84:1

Griffiths et al. Introduction to Genetic Analysis, 10th ed. (2012)

13 / 38

Mendel's first law of equal segregation

Expression in modern terms:

1. A hereditary factor called **gene** is responsible for pea color
2. Each **diploid** plant has a **pair** of this type of gene
3. Each gene occurs in two forms (**alleles**), Y and y
4. A plant is either Y/Y, Y/y or y/y
5. In the Y/y plant the Y allele is **dominant** over y, which is **recessive**.
6. In meiosis, the two alleles segregate equally into eggs and sperm.
7. A single gamete contains only one allele of a gene
8. At fertilization, gametes fuse randomly

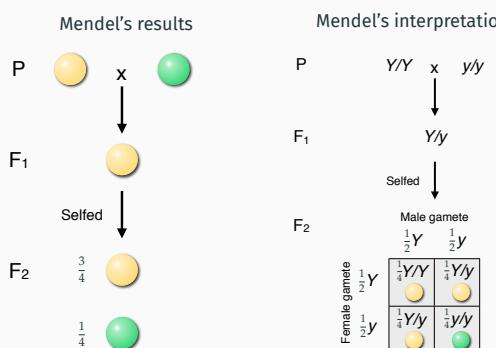
14 / 38

Some terminology

- A fertilized egg is a **zygote**.
- An individual with identical alleles is a **homozygote**
- An individual with two different alleles is a **heterozygote**
- The combination of alleles are called **genotypes**: Y/Y, Y/y or y/y
- Y/Y is **homozygous dominant**
- Y/y is **heterozygous**
- y/y is **homozygous recessive**

15 / 38

Analysis of Mendel's crosses: Selfing



16 / 38

Discussion question 1:



The red arrows show selfing as pollination within single flowers of one F₁ plant.

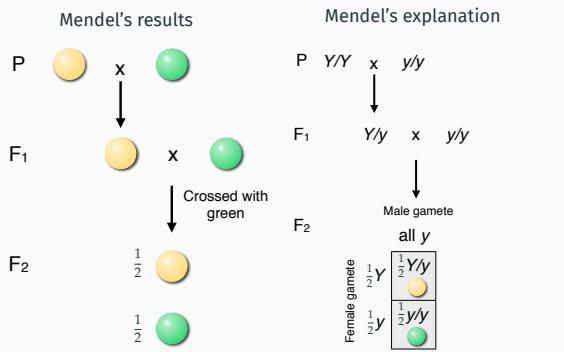
Would the same F₂ results be produced by cross-pollinating two different F₁ plants?

1. Yes.
2. No. (if so, which ratio?)

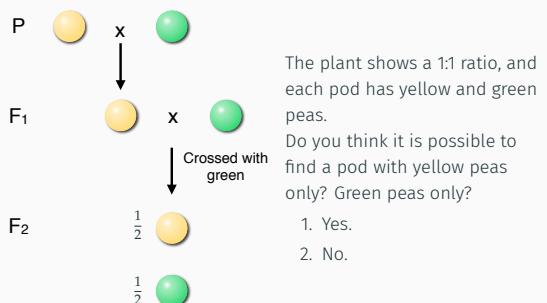
Answer: Yes, because all F₁ plants are heterozygous.

17 / 38

Analysis of Mendel's crosses: Cross-pollination of F₁

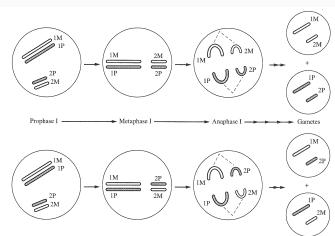


Discussion question 2:



19 / 38

Meiosis and Mendel's law of equal segregation



Equal distribution of gametes on male and female gametes during meiosis.

20 / 38

Background

Mendel's first rule: Equal segregation of variation

Mendel's second rule: Independent Assortment of Traits

21 / 38

Assortment of useful genes into varieties

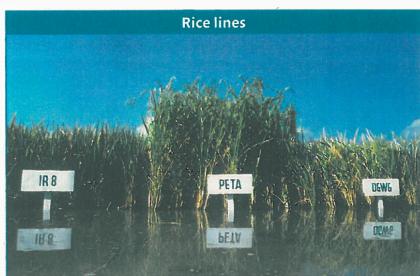


Figure 3-1 Superior genotypes of crops such as rice have revolutionized agriculture. This photograph shows some of the key genotypes used in rice breeding programs. [International Rice Research Institute.]

Griffiths et al. Introduction to Genetic Analysis, 10th ed. (2012)

22 / 38

So far, we have considered the inheritance of single genes. Plant genomes, however, contain tens of thousands of genes. What does this mean for the inheritance patterns of groups of genes? This is an important issue, however, in plant breeding because modern plant varieties contain certain combinations of traits that frequently originated in different plant varieties or geographic regions, but are now contained in single varieties.

This is exactly what happened in the green revolution.

Breeders were looking for traits that originated from single-gene mutations that they could combine to produce a high yielding "Green Revolution" variety.

"Green Revolution" genes in rice

'Positive' alleles of these genes:

- *sd1*: Short stature, less lodging
- *se1*: Independence of photoperiod
- *Xa4*: Resistance against bacterial blight
- *bph2*: Resistance against grasshoppers
- *Snb1*: Resistance against water submersion

Breeding goal: Combine these alleles into a single variety.

23 / 38

- *sd1*: The recessive allele gave the plants a short stature and made them more resistant to lodging
- *se1*: The recessive allele made the plant independent of daylength and allowed it to be cultivated at different latitudes
- *Xa4*: the dominant allele brings resistance against an important disease, bacterial blight
- *bph2*: The allele provides resistance against grasshoppers
- *Snb1*: the dominant alleles cause plants to be resistant against submersion in water after heavy rains

Assortment of genes into varieties

1. Identify mutant lines with desired genes
2. Breeders assort them into desirable combinations
3. Cross individuals with different genes
4. Select offspring that combine favorable traits/genes

24 / 38

Inheritance of multiple genes

- Mendel analysed crosses of parental plants that differed in **two** traits.
- Lines that are homozygous for two traits have genotypes $A/A \cdot B/B$
- The \cdot indicates: We do not know whether the genes are on the same chromosome
- If they are on the same chromosome: AB/AB
- If they are on different chromosomes: $A/A; B/B$

Mendel developed the second rule from crosses of **double heterozygotes** or **dihybrids**:

$$\frac{A}{a} \cdot \frac{B}{b} \times \frac{A}{a} \cdot \frac{B}{b}$$

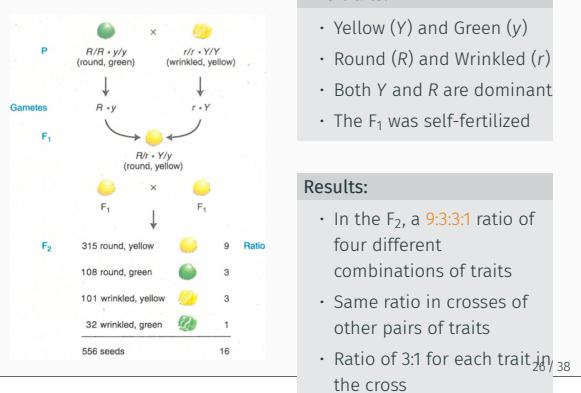
25 / 38

It makes a big difference whether two genes are located on the same chromosome or on different chromosomes.

If they are on different chromosomes, the pairs of chromosomes are independently distributed.

In this case, the alleles of two heterozygous gene pairs show independent assortment.

Mendel's experiment on independent segregation



Two traits:

- Yellow (Y) and Green (y)
- Round (R) and Wrinkled (r)
- Both Y and R are dominant
- The F₁ was self-fertilized

Results:

- In the F₂, a 9:3:3:1 ratio of four different combinations of traits
- Same ratio in crosses of other pairs of traits
- Ratio of 3:1 for each trait in the cross

Mendel obtained dihybrids in the F₁ by crossing lines that differed in two traits.

Mendel used seed shape and seed color

Calculating ratios of phenotypes

The ratios for groups of traits can be calculated from individual traits

$$\frac{3}{4} \times \frac{3}{4} = \frac{9}{16} \quad \text{round, yellow}$$

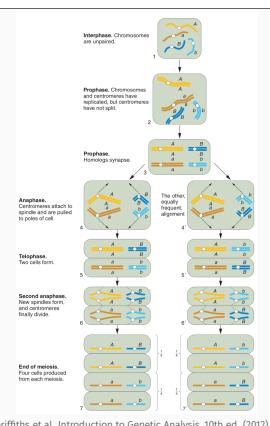
$$\frac{3}{4} \times \frac{1}{4} = \frac{3}{16} \quad \text{round, green}$$

$$\frac{1}{4} \times \frac{3}{4} = \frac{3}{16} \quad \text{wrinkled, yellow}$$

$$\frac{1}{4} \times \frac{1}{4} = \frac{1}{16} \quad \text{wrinkled, green}$$

Biological interpretation: Gene pairs on different chromosomes assort independently during meiosis

27 / 38



Griffiths et al. Introduction to Genetic Analysis, 10th ed. (2012)

28 / 38

Calculate the frequency of gametes

What is the frequency of the different gamete types with round (R) and wrinkled (r) and yellow (Y) and green (y) alleles?

- In double homozygotes there is only 1 type of gametes
- In double heterozygotes there are 4 types of gametes

$\frac{1}{2}$ gametes are R :

- $\frac{1}{2}$ or these R gametes will be Y
- $\frac{1}{2}$ will be y

$\frac{1}{2}$ gametes are r :

- $\frac{1}{2}$ or these r gametes will be Y
- $\frac{1}{2}$ will be y

29 / 38

Frequency of gamete types (continued)

Calculate the frequency of gamete types using the product rule for independent events:

$$R, Y \quad \frac{1}{4}$$

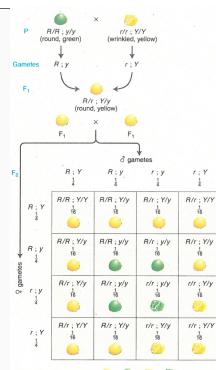
$$R, y \quad \frac{1}{4}$$

$$r, Y \quad \frac{1}{4}$$

$$r, y \quad \frac{1}{4}$$

These gametes mate randomly under the assumption of **random mating**.

30 / 38



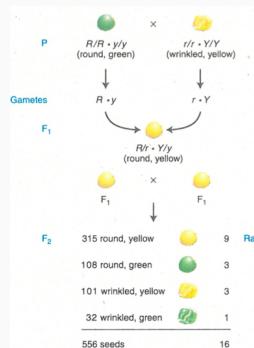
Punnett square

- Model the independent assortment
- Calculate the 9:3:3:1 ratio

Griffiths et al. Introduction to Genetic Analysis, 10th ed. (2012)

31 / 38

Discussion question 3:



If Mendel had selfed the F₁, would he have obtained the same segregating ratios?

1. Yes.
2. No.

32 / 38

Summary

- The two rules of Mendel describe the inheritance of genes
- Mendel used phenotypes controlled by single genes
- Independent and random assortment of chromosomes is the basis of both rules
- Another requirement is random union of gametes
- After the rediscovery of Mendel's laws, the same ratios were observed in many other species!
- Mendel's law is the conceptual basis of combining favorable traits in plant breeding

33 / 38

Further reading

- Original text of Mendel's paper:
http://www.deutschestextarchiv.de/book/view/mendel_pflanzenhybriden_1866
- Griffiths et al., Introduction to Genetic Analysis, 10th Edition, 2012. Chapter 2 "Single Gene Inheritance"

34 / 38

Review questions i

1. What is the difference between a wild type and a mutant?
2. How would you determine if you observe a phenotypic polymorphism, which is the wild type and which the mutant?
3. What is the difference between a genotype and a phenotype?
4. What are the two types of crossing schemes that Mendel used?
5. What are the key characteristics of a Mendelian trait?
6. What are the common features of the seven traits that Mendel studied, and how did this affect his conclusions regarding the inheritance of traits?

35 / 38

Review questions ii

7. How would the segregation ratios be affected if chromosomes do not assort independently during meiosis?
8. Does it make a difference for the segregation ratios if dihybrids are crossed with another dihybrid F₁ plant or if many F₁ plants are selfed?
9. Why did Mendel not have to know much about DNA and genes to reach his conclusions?
10. What is the difference between a wild type and a mutant?
11. How would you determine if you observe a phenotypic polymorphism, which is the wild type and which the mutant?

36 / 38

Review questions iii

12. What is the difference between a genotype and a phenotype?
13. What are the key statements of Mendel's first and second law?
14. Which aspect of plant breeding utilizes Mendel's rules?

37 / 38

References i

-  Krizek, B. A. and Fletcher, J. C. (2005). Molecular mechanisms of flower development: an armchair guide. *Nature Reviews Genetics*, 6(9):688-698.

38 / 38



Molecular Markers and Genetic Maps

3501-211 Genetische Grundlagen der Pflanzenzüchtung

Prof. Karl Schmid

WS 2023/2024

Institute of Plant Breeding, Seed Science and Population Genetics
University of Hohenheim

1 / 46

Overview

Definition of genetic maps

Construction of genetic maps

2 / 46

Recapitulation of the last lectures

Molecular genetics

- The genome consists of DNA that is packed as **chromosomes**
- Chromosomes can be **tightly packed** or unwound
- The genome harbors functional elements (**genes**) and **nonfunctional DNA**
- Gene regulation** and **expression** are complex molecular processes

Mendelian genetics

- Mendel's rules describe the segregation of traits
- Equal segregation** of all alleles onto the gametes of the next generation
- Independent assortment** of chromosomes and their genes

3 / 46

Definition of genetic maps

Construction of genetic maps

4 / 46

The identity and location of Mendel's seven genes/traits

Trait	Dominant phenotype	Recessive phenotype	Symbol group	Linkage group	Cloned	Gene function	Molecular nature of mutation
Seed shape	Round	Wrinkled	R	V	Yes	Starch branching enzyme 1	0.8-kb insertion
Stem length	Tall	Dwarf	LE	III	Yes	GA 3-oxidase	G-to-A substitution
Calyx color	Yellow	Green	I	I	Yes	Day-night gene	Stop mutation
Seed coat/flower color	Purple	White	A	II	Yes	bHLH transcription factor	G-to-A at splice site
Pod color	Green	Yellow	GP	V	No	Chloroplast structure in pod wall	Unknown
Pod form	Inflated	Constricted	d?	III	No	Sclerenchyme formation in pods	Unknown
Position of flowers	Axial	Terminal	FA	IV	No	Monstien function	Unknown

[Reid and Ross, 2011]

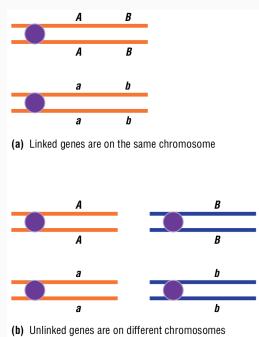
5 / 46

It is noteworthy that almost all of the genes that Mendel studied are located on different chromosomes.

This could either be chance, which is small, however, because the haploid genome of pea consists of only 7 chromosomes, or because Mendel cheated because he selected those traits that *after* his investigation showed the 9:3:1 ratio. We likely will never know as his records of the experiments were lost.

More importantly, his analyses strongly depend on the material he collected from pea varieties that were grown in Eastern Europe at that time and showed variation in a particular set of genes that influence the trait of interest (in other words, a yellow pea color may be caused by another gene located elsewhere in the genome).

What is genetic linkage?



Meneely et al. (2017) Genetics, Oxford University Press

6 / 46

Key questions of interest for geneticists

- What genes are present in the genome?
- What is the function of the genes?
- Where are the genes located in the genome?

The last question is studied with **genetic mapping** and **genome sequencing**

Genetic mapping has important applications in breeding.

7 / 46

Genetic mapping has two applications in breeding:

- It allows to design crosses aimed at combining favorable alleles of different genes onto a single individual. For example in resistance breedings different alleles that provide resistance in different resistant genes are combined in a single genotype to provide a durable resistance.
- The length and the rates of recombination on a chromosome are defined in a genetic map and allow to characterize the rate of recombination in a genetic region.

Two types of maps in genetics

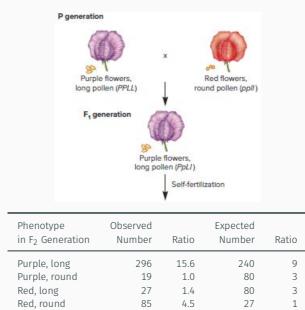


- Physical map: Distance between genes in **basepairs**
- Genetic map: Distance between genes in **genetic map units** (**centiMorgan**)

8 / 46

Genetic linkage

Genetic linkage was discovered by the geneticists William Bateson and Reginald Punnett in 1908



9 / 46

The recombination frequency between two genes (or markers) is expressed as percent of offspring that shows a recombination event. The recombination frequency is then associated with the genetic distance between genes based on the notion that a higher distance between genes increases the probability of a recombination event.

Genetic linkage maps

- Recombination frequency between genes in percent
- 1 genetic map unit (m.u.) = 1/100 products out of meiosis are recombinant
- 1 m.u. = 1 centimorgan (cM)
- This measure is named after Thomas Morgan (Drosophila geneticist)



Thomas Hunt
Morgan
(1866-1945)
Nobelpreis 1933

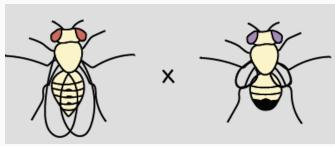
10 / 46

Why are genetic maps useful?

- Position of genes required to build complex genotypes by breeding and selection
→ **Important in plant breeding!**
- Develop genetic markers linked to useful alleles and use markers to select offspring
→ **Marker assisted selection (MAS)**
- Identify ("map") and study the function of important genes
- Analyse the genetic history and diversity of genes
→ e.g. to identify new useful genetic variation for plant breeding

11 / 46

Analysis of genetic linkage in Drosophila



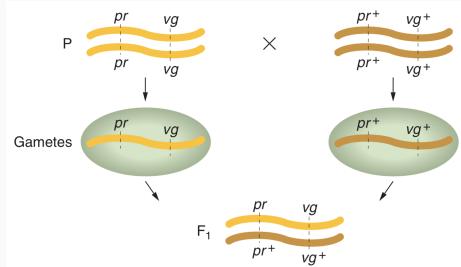
CCO / Public domain

Two Mendelian genes visible as phenotypes

- pr: purple vs. red eyes
- vg: vestigial - short vs. long wings

12 / 46

Linked alleles tend to be inherited together...

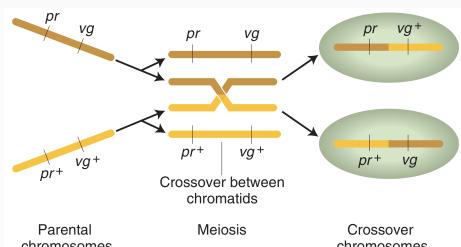


Griffiths et al. Introduction to Genetic Analysis, 10th ed. (2012)

Simple inheritance of two traits: The same alleles of two genes from the parent and offspring are on the same chromosome

13 / 46

... but recombination produces new combinations of alleles



Griffiths et al. Introduction to Genetic Analysis, 10th ed. (2012)

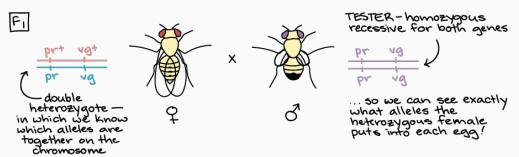
14 / 46

Linkage between genes: A cross by T. H. Morgan

P $\frac{pr}{pr} \cdot \frac{vg}{vg}$ \times $\frac{pr+}{pr+} \cdot \frac{vg+}{vg+}$
 ↓
 Gametes $pr \cdot vg$ $pr+ \cdot vg+$
 ↓
 F₁ dihybrid $\frac{pr+}{pr} \cdot \frac{vg+}{vg}$
 Testcross $\frac{pr+}{pr} \cdot \frac{vg+}{vg}$ \times $\frac{pr}{pr} \cdot \frac{vg}{vg}$
 F₁ dihybrid female Tester male

15 / 46

Linkage between genes: A cross by T. H. Morgan



CCO - Public domain

16 / 46

Linkage between genes: A cross by T. H. Morgan

P $\frac{pr}{pr} \cdot \frac{vg}{vg}$ \times $\frac{pr+}{pr+} \cdot \frac{vg+}{vg+}$
 ↓
 Gametes $pr \cdot vg$ $pr+ \cdot vg+$
 ↓
 F₁ dihybrid $\frac{pr+}{pr} \cdot \frac{vg+}{vg}$
 Testcross $\frac{pr+}{pr} \cdot \frac{vg+}{vg}$ \times $\frac{pr}{pr} \cdot \frac{vg}{vg}$
 F₁ dihybrid female Tester male

Result:	
$pr+ \cdot vg+$	1339
$pr \cdot vg$	1195
$pr+ \cdot vg$	151
$pr \cdot vg+$	154
	2839

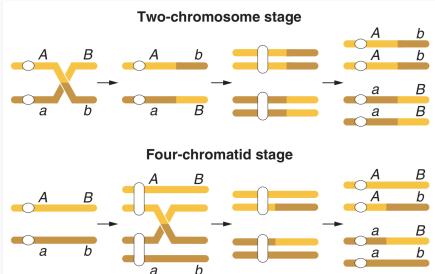
17 / 46

pr stands for purple / violet
 vg stands for vestigial - small wings

Question: What is the expectation of the numbers if the genes are located on different chromosomes?
 A frequency of 710 on each of the gametes.

Recombination occurs after chromosome replication

Crossing over is between chromatids, not chromosomes



Griffiths et al. Introduction to Genetic Analysis, 10th ed. (2012)

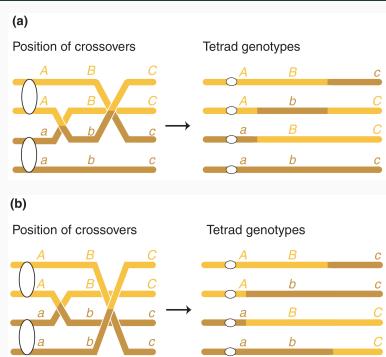
18 / 46

- Exchange of parts of the chromosome by crossing over produces new genetic variations that differ from the parent.

- The analysis of tetrads (= products of meiosis) shows that recombination occurs **after replication** in the four chromatid stage.

Tetrads can be investigated easily in model organisms like the fungus *Neurospora crassa*, where the cells originating from meiosis remain attached together in the same order of their origin.

Multiple crossovers can involve more than two chromatids



Griffiths et al. Introduction to Genetic Analysis, 10th ed. (2012)

19 / 46

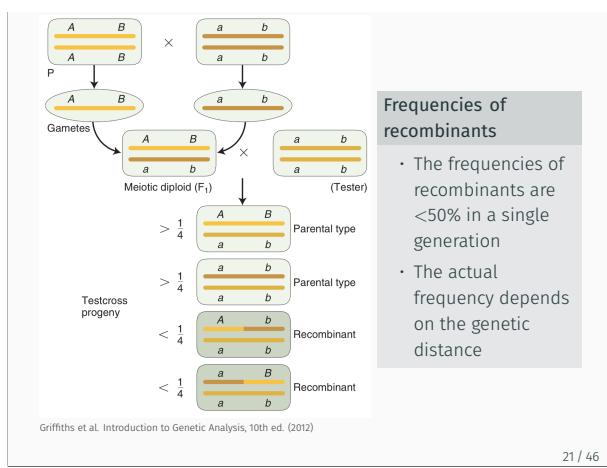
Tetrad analysis that compare three different loci showed that crossovers can occur over multiple chromatids.

Recombinants are produced by crossovers

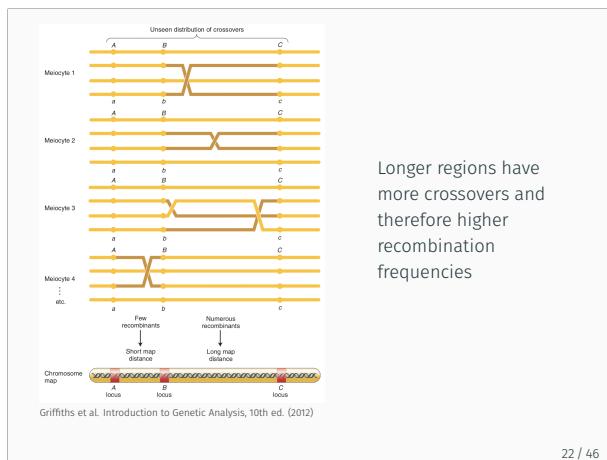
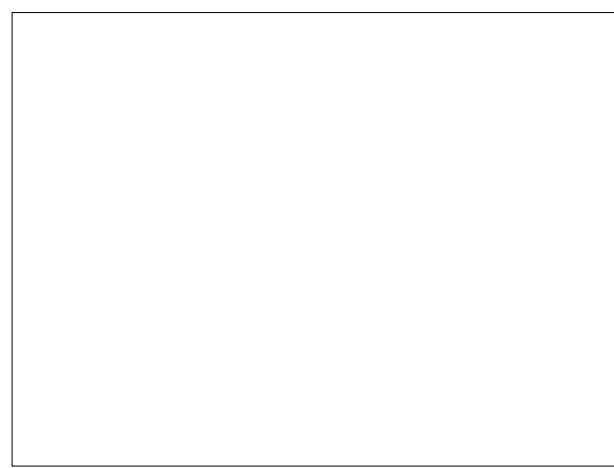
	Meiotic chromosomes	Meiotic products	
Meioses with no crossover between the genes			Parental Parental Parental Parental
Meioses with a crossover between the genes			Parental Recombinant Recombinant Parental

Griffiths et al. Introduction to Genetic Analysis, 10th ed. (2012)

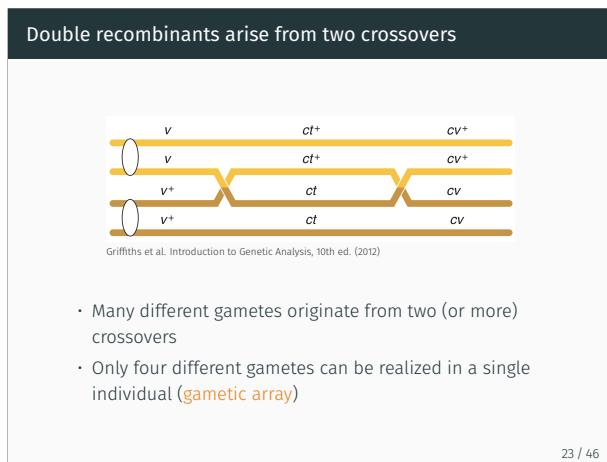
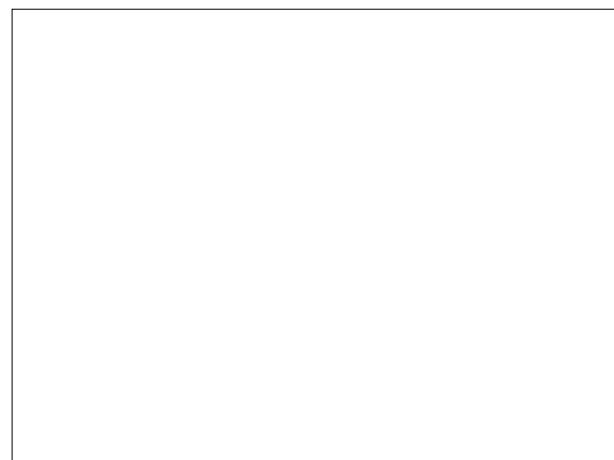
20 / 46



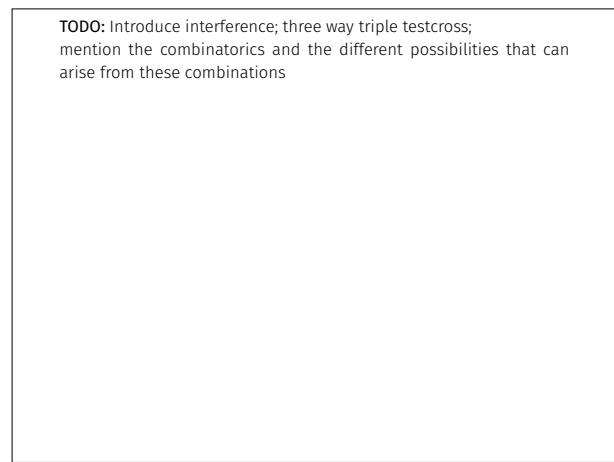
21 / 46



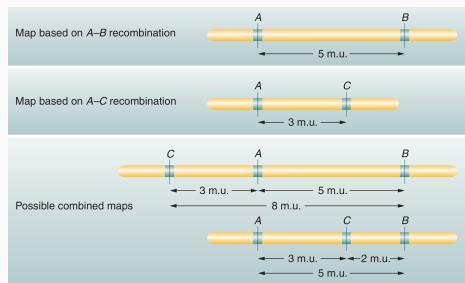
22 / 46



23 / 46



Map distances are generally additive



24 / 46

Definition of genetic maps

Construction of genetic maps

25 / 46

Construction of a genetic map: Qualitative traits

Which traits are useful for constructing a genetic map?

- Should be **Mendelian traits**: Two states representing two different alleles
- Should be easy to score
- Should be located on the same chromosome

In the 'old' days, phenotypes were used, today **genetic markers** are mainly used.

26 / 46

Locus: *shrunken 1* (*sh1*)

Phenotype: Collapsed endosperm - shrunken



27 / 46

Phenotype of *shrunken 1* (*sh1*)

- Kernel:
 - Normal: *Sh1*
 - shrunken: *sh1*
- Gene product: Sucrose synthase 1 (*SUS-SH1*)
- Linkage group: 9
- Phenotype: shrunken kernel, reduced starch level, early cell degeneration

28 / 46

Locus: *waxy 1* (*wx1*)

Phenotype: amylopectin red

Kernel



Pollen segregation

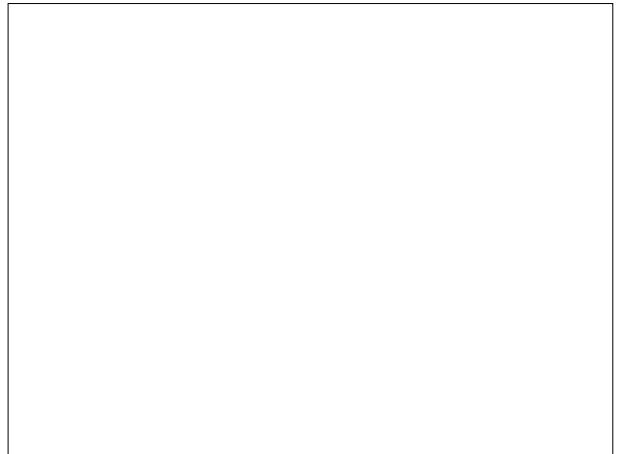


29 / 46

waxy 1 (*wx1*)

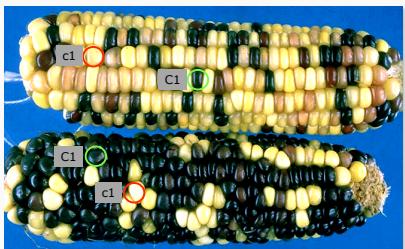
- Starch (endosperm & pollen)
 - Amylose (blue with iodine) *Wx1*
 - Amylopectin (red with iodine) *wx1*
- Gene product: starch-granule-bound nucleotide diphosphate-starch glycosyl transferase
- Linkage group: 9
- Phenotype: red colored starch with iodine

30 / 46

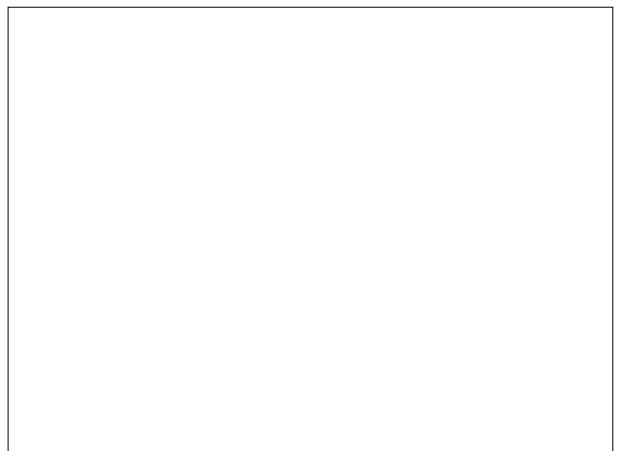


Locus: colored aleurone 1 (*c1*)

Phenotype: Colored aleuron



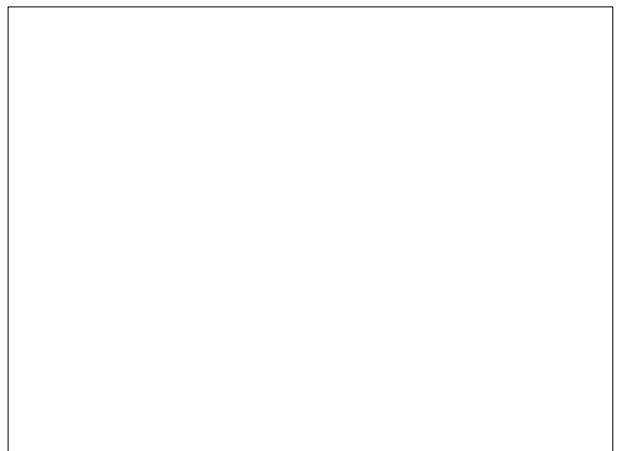
31 / 46



colored aleurone 1 (*c1*)

- Aleuron
 - Colored *C1*
 - colorless *c1*
- Gene product: dominant colored, anthocyanin biosynthesis
- Linkage group: 9
- Phenotype: colorless aleurone

32 / 46



Construction of a linkage map in maize

Parent 1 Phenotype: Colored shrunken Blue

Genotype: $\frac{C\ sh\ Wx}{C\ sh\ Wx}$ or $\frac{+ sh +}{+ sh +}$

Parent 2 Phenotype: Colorless normal red

Genotype: $\frac{c\ Sh\ wx}{c\ Sh\ wx}$ or $\frac{c + wx}{c + wx}$

Note: + indicates the wild type (non-mutant) allele.

33 / 46

Recombination frequencies

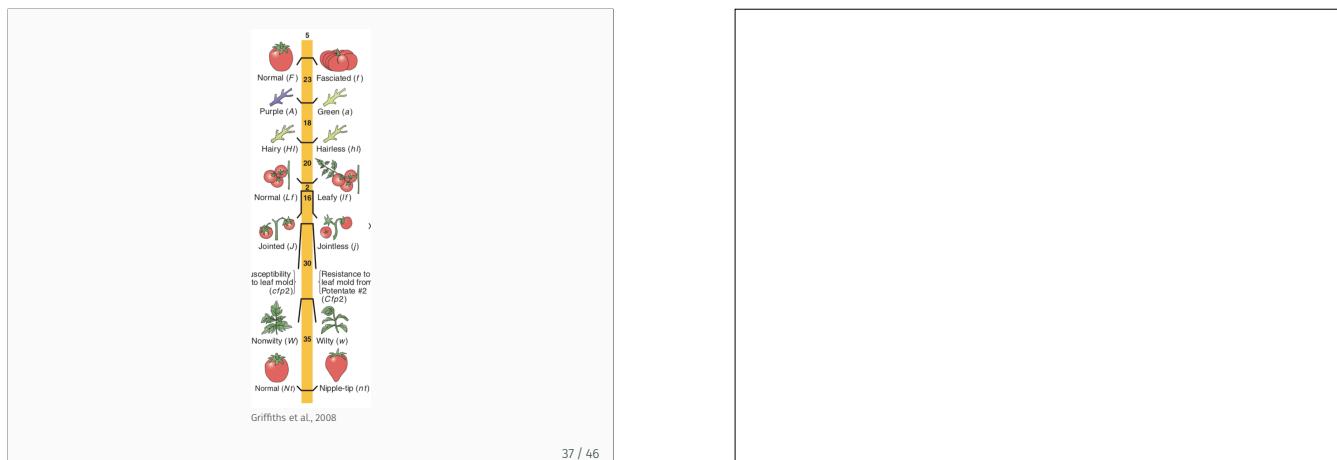
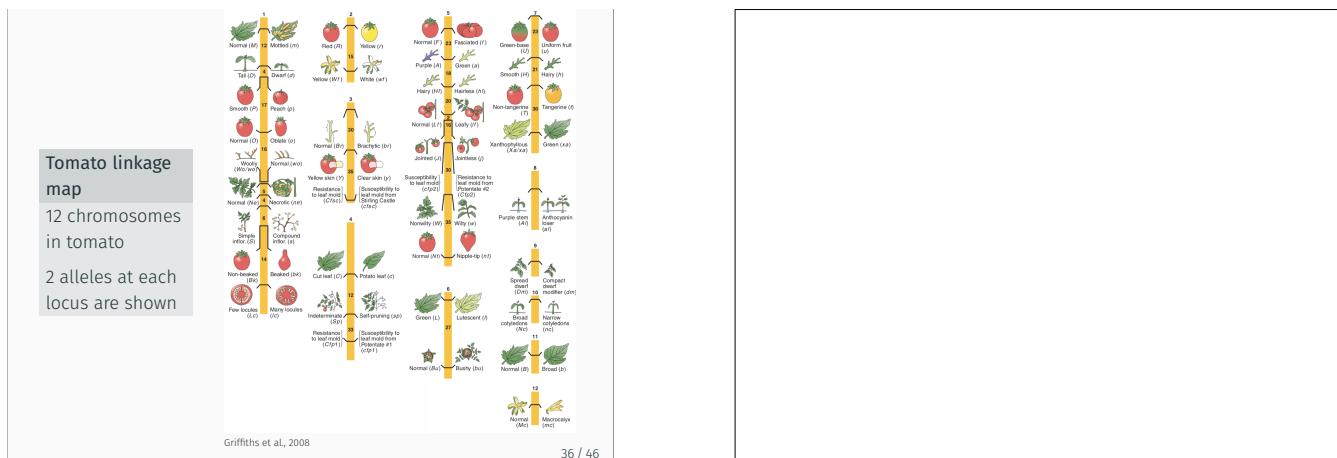
	Phenotypes	Genotypes	Recombinants			Sum
			c-sh	sh-wx	c-wx	
P1	colored shrunken blue	+ sh +	0	0	0	2538
	colored shrunken blue	c sh +	0	0	0	2708
P2	colorless normal red	c + wk	0	0	0	2708
	colorless normal red	c sh +	0	0	0	2708
F2	colored normal red	+ + wx	116	0	116	116
	colorless shrunken blue	c sh +	113	0	113	113
	colored shrunken red	+ sh wx	0	602	602	602
	colorless normal blue	c + +	0	626	626	626
	colored normal blue	+ + +	4	4	0	4
	colorless shrunken red	c sh wx	2	2	0	2
			235	1234	1457	6709

34 / 46

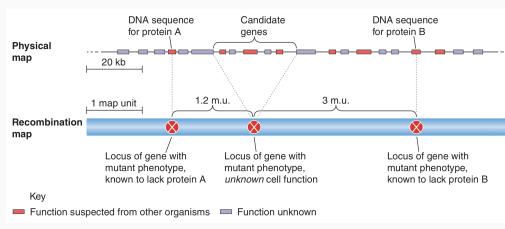
Generation of a linkage map

Recombinants	c - sh	sh - wx	c - wx
Total	235 /	1234 /	1457 /
6709	6709	6709	6709
	3.5%	18.4%	21.7%

35 / 46



Comparison of physical and genetic maps

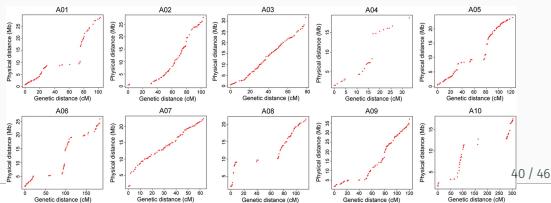


39 / 46

Comparison of genetic and physical maps in Chinese cabbage



Source: [Huang et al., 2017]



Source: [Huang et al., 2017]

Summary: Construction of a genetic map

- Physical maps measure the physical distance between loci
- Genetic (= Linkage) maps are based on recombination
- Genetic maps show the recombination frequency between markers
- Definition of recombination frequency:
Map units = centimorgan

41 / 46

Further reading

- Chapter 4 in [Griffiths, Antony J. F. et al., 2012]
- Chapter 9 in [Meneely et al., 2017]
- <https://www.khanacademy.org/science/biology/classical-genetics/chromosomal-basis-of-genetics/a/linkage-mapping>

42 / 46

Review questions i

1. What are the key differences between a genetic and a physical map?
2. Define the terms 'parental' and 'recombinant' in the context of genetic linkage.
3. What is the relationship between the rate of recombination and the length of a genetic map?
4. Why do the segregation frequencies of linked genes differ from the segregation frequencies of unlinked genes (genes on different chromosomes)?
5. How is it possible that multiple chromatids are involved in recombination?

43 / 46

Review questions ii

6. How is the genetic distance of two or more genes calculated from the offspring of a cross?
7. Why are the genetic distances of genes on a chromosome generally additive?
8. What are the key steps in the creation of a genetic map?
9. How can genetic and physical maps complement each other in genetic experiments if the goal is to identify genes controlling a trait?
10. What are the uses of a genetic and physical map for plant breeding?

44 / 46

References i

- [pdf] Griffiths, Antony J. F., Wessler, Susan R., Carroll, Sean B., and Doebley, John (2012). *Introduction to Genetic Analysis*. W. H. Freeman and Company, 10th edition.
- [pdf] Huang, L., Yang, Y., Zhang, F., and Cao, J. (2017). A genome-wide SNP-based genetic map and QTL mapping for agronomic traits in Chinese cabbage. *Scientific Reports*, 7(1):46305. Bandiera_abtest: a Cc_license_type: cc_by Cg_type: Nature Research Journals Number: 1 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Agricultural genetics;Plant breeding Subject_term_id: agricultural-genetics;plant-breeding.
- [pdf] Meneely, P. M., Hoang, R. D., Okeke, I. N., and Heston, K. (2017). *Genetics: genes, genomes, and evolution*. Oxford University Press.
- [pdf] Reid, J. B. and Ross, J. J. (2011). Mendel2019s Genes: Toward a Full Molecular Characterization. *Genetics*, 189(1):3–10.

45 / 46

References ii

- [pdf] Ren, X., Wang, J., Liu, L., Sun, G., Li, C., Luo, H., and Sun, D. (2016). SNP-based high density genetic map and mapping of btwd1 dwarfing gene in barley. *Scientific Reports*, 6:31741.

46 / 46



UNIVERSITY OF
HOHENHEIM

Genetic Markers:

Definition, analysis and quantification

3501-211 Genetische Grundlagen der Pflanzenzüchtung

Prof. Karl Schmid
WS 2023/2024

Institute of Plant Breeding, Seed Science and Population Genetics
University of Hohenheim

1 / 31

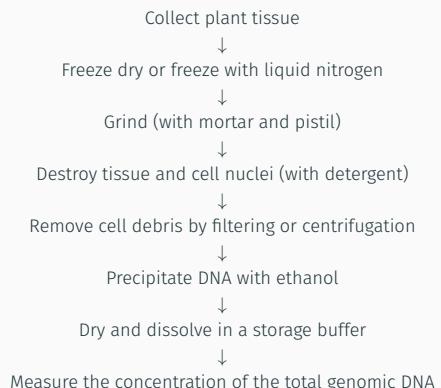
Genetic diversity analysis in plants

Steps in the analysis of genetic variation:

1. Collecting and listing of germplasm
2. Plant cultivation
3. DNA isolation
4. DNA marker analysis
5. Data analysis and interpretation

2 / 31

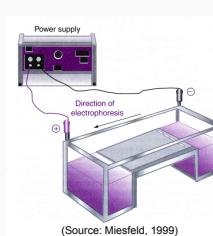
The isolation of DNA from tissue



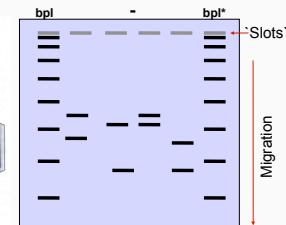
3 / 31

Analysis of DNA

Gel electrophoresis: separation of molecular markers by electrophoretic mobility



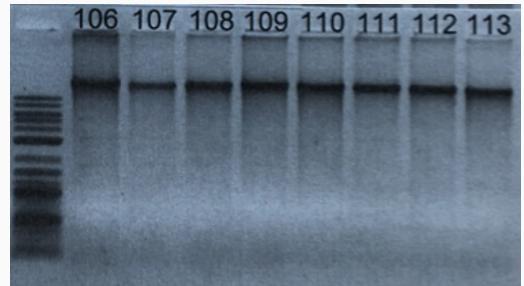
(Source: Miesfeld, 1999)



*bpl = base pair ladder: reference for estimating the size of fragments

4 / 31

Example of an agarose gel



5 / 31

Introduction

Genetic markers

Quantification of Genetic Variation

Summary

6 / 31

Quantification of genetic variation

- Goal: Define a way to measure the level of genetic variation
- Analyse diversity using different marker types
- Apply measures of genetic variation in the context of genetic resources

7 / 31

Types of genetic variation

- 1900's: Visible polymorphisms
- 1930's: Chromosomal polymorphisms
- 1940's: Blood groups
- 1960's: Protein polymorphisms
- 1980's: DNA Sequencing of genes
- 2000's: Resequencing of genomes

8 / 31

Types of DNA sequence variation

- Single nucleotide polymorphisms (SNPs)
- Variation in other repetitive elements such as minisatellites or gene families (Copy number variants, CNVs)

9 / 31

Mutations versus Polymorphisms versus Markers

- **Mutation:** The process that generates variation
- **Polymorphism:** A segregating mutation
- **Marker:** A defined polymorphism that can be detected by a marker system, whose location is known in the genome

10 / 31

Frequency of polymorphism types



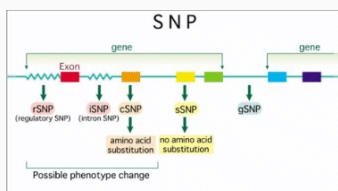
Genome of J. Craig Venter:

- 3.2 Million SNPs
- 0.9 Million CNV and repeat polymorphisms
- SNPs constitute only 26% of the 12.3 Mb of polymorphic DNA
- The remaining variants are insertions and deletions

11 / 31

Classification of SNPs

- **Noncoding:** Located in regions that do not code for proteins
- **Coding:** Located in protein-coding regions
- **Replacement polymorphism:** Replace amino acid
- **Synonymous polymorphism:** Do not replace amino acid
- **Regulatory polymorphism:** include replacement and synonymous polymorphism; may affect gene function



12 / 31

<http://snp.ims.u-tokyo.ac.jp/samplesMethods.html>

Resequencing of complete genomes



Genome-wide association studies of 14 agronomic traits in rice landraces

Kunhai Huang^{1,2*}, Xuguang Wang^{1,3†}, Tao Song^{1,4}, Qiang Zhou^{1,5,6}, Qi Feng^{1,6}, Yan Zhou¹, Jianying Li¹, Chuanming Zou¹, Tingting Lei¹, Zhixia Zhang¹, Meng Li^{1,2}, Daolin Hu¹, Yanchi Guo¹, Abing Wang¹, Le Wang¹, Laiwei Dong¹, Wenzhao Li¹, Yiqi Liu¹, Qian Wang¹, Kunyan Liu¹, Tao Huang¹, Taiyong Zhou¹, Yufeng Jing¹, Wei Li¹, Zhang Lin¹, Edward S Buckler², Qian Qian¹, Qi-Fa Zhang¹, Jiayang Li^{1,6} & Bin Han^{1,2}

Uncovering the genetic basis of agronomic traits in crop landraces that have adapted to various agro-climatic conditions is important for breeding. We have identified 1.1 million SNPs by resequencing 217 rice landraces and constructed a high-density haplotype map of the rice genome using a novel data-imputation method. We performed genome-wide association studies (GWAS) for 14 agronomic traits in the population of Oryza sativa indica subspecies. We identified 166 significant SNPs associated with 14 traits, and found 111 SNPs are located in 100 candidate genes. This study provides a fundamental resource for rice genetics research and breeding, and demonstrates that an approach combining GWAS and QTL analysis based on the haplotype map of rice landraces can be used as a powerful complementary strategy to classical biparental cross-mapping for dissecting complex traits in rice.

- Ongoing projects in many plant crops (Rice, maize, sorghum, etc.)

13 / 31

Introduction

Genetic markers

Quantification of Genetic Variation

Summary

14 / 31

Analysis of genetic diversity: Objectives

- **Analysis of genetic diversity in germplasm**
- **Management and organisation of gene banks:** Identification of duplicates, definition of core collections
- **Parameters of genetic diversity:** Extent of variation within and between populations
- **Evolutionary studies:** Phylogenetic relationships
- **Taxonomic studies:** Species identification
- **Plant breeding:** Selection, prediction of breeding values

15 / 31

Some definitions

- **Locus:** Any region in a genome (e.g., a gene)
- **Allele:** The state (i.e., sequence) of a single chromosome in a region
- **Genotype:** The combination of two alleles (in a diploid organism) on an individual in a region

16 / 31

Genotype and allele frequencies

- Locus A, B, C, ...
- Alleles at locus A: A₁, A₂, ...
- Frequency of alleles: A₁ : p, A₂ : q

Genotype:	A ₁ A ₁	A ₁ A ₂	A ₂ A ₂
Relative frequency:	x ₁₁	x ₁₂	x ₂₂

Sum of relative frequencies: x₁₁ + x₁₂ + x₂₂ = 1

17 / 31

Simple measures of genetic variation

- Polymorphisms: Fraction of polymorphic loci

$$P = \frac{\text{number of polymorphic loci}}{\text{number of all investigated loci}}$$

- Heterozygosity with n loci:

$$H_{obs} = \frac{1}{n} \sum_{i=1}^n H_i,$$

18 / 31

Measures of genetic distance

- Genetic similarity is the proportion of marker states that are shared between two samples.
- Many different modifications of this simple definition
- Applies to all types of markers

Example:

Sequence 1 ACTGTGCTGA
Sequence 2G..
Sequence 3 T.A....T.A.
Sequence 4 T.A.G.T.A.

Note: A " indicates identical nucleotide to Sequence 1.

19 / 31

Genetic similarity of 6 maize inbred lines

- The results are based on DNA sequence data (100 base pairs)

Line	2	3	4	5	6
1	0.78	0.67	0.60	0.54	0.52
2		0.54	0.69	0.64	0.53
3			0.59	0.49	0.51
4				0.70	0.60
5					0.60

- Highest genetic similarity between lines 1 and 2
- Lowest genetic similarity between lines 3 and 5

20 / 31

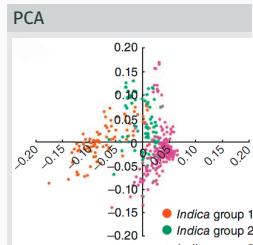
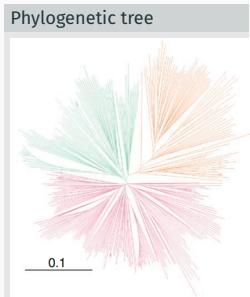
Clustering and ordination methods

- Graphical representation of marker data
- Simplify complex data sets
- Identify relationships in the data
- Two classes of methods: Ordination methods and clustering methods
- Ordination method: Principal components analysis
- Clustering methods: Phylogenetic trees
- Numerous variants of both methods were developed

21 / 31

Genome resequencing of landraces of rice

- 373 *indica* landraces from China were resequenced
- Clustering by phylogenetic tree or PCA

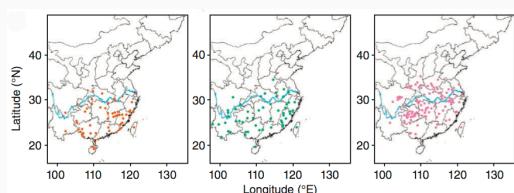


22 / 31

Source: [Huang et al., 2010]

Genome resequencing of rice landraces

- Transfer the three clusters onto a map using information on geographic origin
- There is genetic clustering, but no geographic clustering of the three main groups of *indica* landraces



23 / 31

Quantification of DNA sequence variation

A DNA sequence alignment:

Sequence 1 ACTGTGCTGA
Sequence 2G..
Sequence 3 T.A...T.A.
Sequence 4 T.A.G.T.A.

Note: A " indicates identical nucleotide to Sequence 1.

- Nucleotide polymorphism:

$$P_n = \frac{n_p}{n_t}$$

with n_p as the number of polymorphic nucleotide positions and n_t as the total number of sequenced nucleotide polymorphisms

24 / 31

Quantification of DNA sequence variation

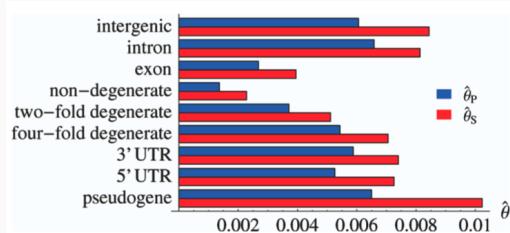
A frequently used measure of DNA sequence variation is **nucleotide diversity**, which is defined as average nucleotide pairwise divergence:

$$\pi = \frac{n}{n-1} \sum_{i=1}^k \sum_{j=1}^k p_i p_j \pi_{ij}$$

- n : number of individuals in sample
- k : number of haplotypes in sample
- p_i : proportion of haplotype i
- π_{ij} : the pairwise differences of haplotypes i and j , measured as the proportion of different SNPs between two haplotypes.

25 / 31

Polymorphism levels in different sites in *Arabidopsis thaliana*



Two estimates: θ_p uses average of pairwise differences ($= \pi$), θ_s : uses the number of polymorphic sites

Source: [Nordborg et al., 2005]

26 / 31

Summary

- DNA markers are genetic polymorphisms that are used to estimate genetic diversity in populations.
- Until recently, only few markers could be technically analysed, but now whole-genome sequencing allows to detect essentially all polymorphisms in a genome.
- Markers are important for a wide range of applications in plant breeding.

27 / 31

Literature

- George Acquaah: Principles of Plant Genetics and Breeding, Blackwell Publishing (2007) - A very good introduction into the topic.
- Griffiths et al: Modern Genetic Analysis, Freeman, New York

28 / 31

Review Questions i

- What are the important steps in the extraction of DNA from living tissues?
- What are the differences between a mutation, a polymorphism and a marker?
- What is the difference between an allele and a genotype?
- How are allele and genotype frequencies calculated?
- What are the applications in the study of genetic variation?
- Which measures are used to measure the level of genetic variation?
- Which measure is available to measure the genetic distance?

29 / 31

Review Questions ii

- Which methods are available to reduce the complexity of genetic datasets and how do they work?
- Which measures are available to quantify the level of DNA sequence variation?

30 / 31

References i

-  Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., Li, C., Zhu, C., Lu, T., Zhang, Z., Li, M., Fan, D., Guo, Y., Wang, A., Wang, L., Deng, L., Li, W., Lu, Y., Weng, Q., Liu, K., Huang, T., Zhou, T., Jing, Y., Li, W., Lin, Z., Buckler, E. S., Qian, Q., Zhang, Q.-F., Li, J., and Han, B. (2010). Genome-wide association studies of 14 agronomic traits in rice landraces. *Nature Genetics*, 42(11):961–967.
-  Nordborg, M., Hu, T. T., Ishino, Y., Jhaveri, J., Toomajian, C., Zheng, H., Bakker, E., Calabrese, P., Gladstone, J., Goyal, R., Jakobsson, M., Kim, S., Morozov, Y., Padhukasasnam, B., Plagnol, V., Rosenberg, N. A., Shah, C., Wall, J. D., Wang, J., Zhao, K., Kalbfleisch, T., Schulz, V., Kreitman, M., and Bergelson, J. (2005). The pattern of polymorphism in *Arabidopsis thaliana*. *PLoS Biol*, 3(7):e196.

31 / 31



Population Genetics - Part 1

3501-211 Genetische Grundlagen der Pflanzenzüchtung

Prof. Karl Schmid
WS 2023/2024

Institute of Plant Breeding, Seed Science and Population Genetics
University of Hohenheim

1 / 42

Overview of lectures on population and quantitative genetics

Population Genetics

- The importance of genetic variation
- The Hardy-Weinberg Equilibrium
- Inbreeding
- Genetic drift
- Selection
- Recombination and Linkage

Quantitative Genetics

- Basic model of quantitative genetics
- Variance components and their estimation
- Phenotypic selection

2 / 42

Outline

- Introduction
- What is genetic variation?
- Quantification of genetic variation
- Hardy-Weinberg Equilibrium

3 / 42

Introduction

What is genetic variation?

Quantification of genetic variation

Hardy-Weinberg Equilibrium

4 / 42

Why study genetic variation?

A collection of Peruvian maize landraces.



5 / 42

This picture shows a basket with peruvian maize landraces.
The differences between the maize cobs result from genetic variation

Landessortenversuch - Baar (Donaueschingen)



6 / 42

In this variety recommendation trial different varieties of spelt wheat are cultivated.

The different "look" of these line varieties reflect genetic variation between varieties. Some of these differences are used as traits to reproducibly differentiate between varieties.

Population Genetics and Plant Breeding

Key questions for plant breeders:

- How much **genetic variation** is in my breeding pool?
- How much does my breeding program **reduce variation**?
- How can I efficiently **introduce** new variation?
- How are yield (and other traits) affected by **inbreeding**?
- How do I **analyse and quantify** genetic variation in crops?

7 / 42

Variation is the basis of natural and artificial selection

Charles Darwin: On the Origin of Species (1859)



- Variation between individuals
→ **Population genetics**
- Inheritance of traits
→ **Quantitative genetics**
- Selection by differential reproduction
→ **Selection methods**

8 / 42

Plant breeding can be seen as an analogous process than natural selection in natural populations.

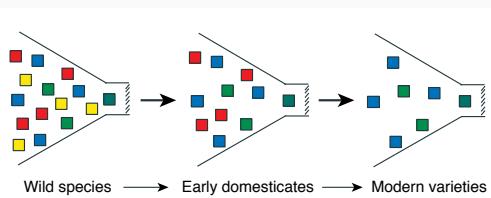
It therefore is no accident that Charles Darwin began his book on the origin of species, which founded evolutionary theory with a chapter on domestication.

Why is it important to study genetic variation?

- Increase genetic diversity in plant breeding
- Identify useful genes for breeding
- Create varieties to particular cultivation conditions (e.g., low- and high input)
- Prepare for climate change
- Create varieties for particular uses (e.g., bioenergy maize)

9 / 42

Loss of genetic variation during domestication and plant breeding

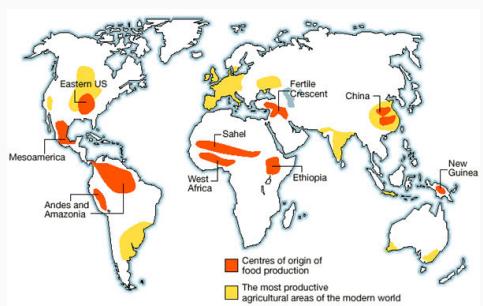


Tanksley and McCouch, Science (1997)

10 / 42

The general theme of this figure is: Domestication and modern plant breeding caused and still cause the loss of genetic diversity.
An important question is whether this is really true and to what extent this is a problem.

Centers of genetic diversity



Diamond, Nature (2002)

11 / 42

The idea that domestication entailed a loss of genetic variation is based on the notion that for most crops, the main areas of cultivation are outside the original center of domestication and that much useful genetic variation was 'left behind' in the centers of domestication.

You will hear more about this in our PGR module

Example: Drought-resistant genotypes of wild barley

Two genotypes of wild barley from the Ein Prat Canyon in Palestine:

- **Desert type:** growing on hot and dry South facing slope is drought and heat tolerant
- **Mediterranean type:** growing on the cooler and moister North facing slope produces more seeds



South facing slope



North facing slope

12 / 42

Barley is in competition with other crops, but the demand for malting barley is increasing.

For this reason, barley is cultivated at extreme sites.

Genetic resources can be used to adapt barley to such sites and therefore make the crop still competitive for the market.

Wild barley at the dead sea



13 / 42

Utilization of genetic diversity in breeding



Backcross into elite variety → Doubled Haploid lines (75% elite, 25% wild)

14 / 42

Outline

Introduction

What is genetic variation?

Quantification of genetic variation

Hardy-Weinberg Equilibrium

15 / 42

Genotypes and phenotypes

Trait: Glume color of barley

Phenotype:



Genotype: bb BB and Bb

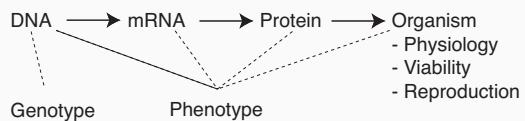
Gene: Glume color: B

Alleles: B : black, b : non-black

16 / 42

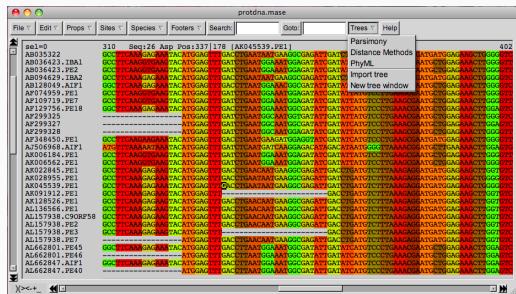
Genetic variation is responsible for phenotypic variation

The 'Central Dogma of Molecular Biology' describes the flow of information from genotype to phenotype:



17 / 42

Sequence variation: A multiple sequence alignment



18 / 42

State of the art: Resequencing of complete genomes

Now it is possible to resequence complete plant genomes and to discover essentially all polymorphisms present in a population:

Genome-wide association studies of 14 agronomic traits in rice landraces

Xuehai Huang^{1,10}, Xinghua Wei^{1,10}, Tao Sang^{1,10}, Qiang Zhao^{1,2,10}, Qi Feng^{1,10}, Yan Zhao¹, Canyang Li¹, Chuangrang Zhu¹, Tingting Lu¹, Zhiwei Zhang¹, Meng Li^{1,6}, Danlin Fan¹, Yunli Guo¹, Ahong Wang¹, Lu Wang¹, Liuwei Deng¹, Wenjun Li¹, Yiqi Liu¹, Oijun Weng¹, Kunyan Liu¹, Tao Huang¹, Taoying Zhou¹, Yufeng Jing¹, Wei Li¹, Zhang Lin¹, Edward S Buckler^{5,7}, Qian Qian³, Qi-Fa Zhang⁴, Jiayang Li³ & Bin Han^{1,2}

Uncovering the genetic basis of agronomic traits in crop landraces that have adapted to various agro-climatic conditions is important to world food security. Here we have identified ~3.6 million SNPs by resequencing 317 rice landraces and constructed a high-density haplotype map of the rice genome using a novel data-imputation method. We performed genome-wide association studies (GWAS) for 14 agronomic traits in the population of *Oryza sativa indica* subspecies. The loci identified through GWAS explained ~36% of the phenotypic variance, on average. The peak signals at six loci were tied closely to previously identified genes. This study provides a fundamental resource for rice genetics research and breeding, and demonstrates that an approach integrating second-generation genome sequencing and GWAS can be used as a powerful complementary strategy to classical biparental cross-mapping for dissecting complex traits in rice.

[Huang et al., 2010]

See also Gigasciences: the 3000 rice genomes project

Outline

Introduction

What is genetic variation?

Quantification of genetic variation

Hardy-Weinberg Equilibrium

20 / 42

Genotype and allele frequencies

Goal: Develop a framework to quantify the level and type of genetic variation

- Locus A, B, C, ...
- Alleles at locus A: A_1, A_2, \dots
- Frequency of alleles: $A_1 : p, A_2 : q$

Genotype: A_1A_1 A_1A_2 A_2A_2
Relative frequency: P H Q

Sum of relative frequencies: $P + H + Q = 1$.

21 / 42

Quantifying population parameters

Genotypes		Alleles	
DD	dd	Dd	dd
DD	dd	DD	dd
dd	Dd	Dd	dd
Dd	DD	dd	Dd
Dd	dd	Dd	DD

Genotype frequencies		Allele frequencies	
Composition	Relative freq.	Composition	Relative freq.
4 Individuals: DD	0.16 = P	20 Alleles D	0.40 = p
12 Individuals: Dd	0.48 = H	30 Alleles d	0.60 = q
9 Individuals: dd	0.36 = Q		
25 Total	1.00	50 Total	1.00

22 / 42

Relationship between genotype and allele frequencies

The frequency p of allele A_1 is

$$p = P + \frac{1}{2}H, \quad (1)$$

and the frequency q of allele a is

$$q = 1 - p = Q + \frac{1}{2}H.$$

The frequency p of allele A_1 has two important characteristics:

- It describes the **relative frequency** of allele A_1 in the population.
- It gives the **probability** that a randomly chosen allele from the population is allele A_1 .

23 / 42

Example calculation: Genotype and allele frequencies

Genotype	Plants	Genotype frequency	Allele	Allele frequency
A_1A_1	100	$P = 0.50$	A_1	$p = P + \frac{1}{2}H = 0.60$
A_1A_2	40	$H = 0.20$	A_2	$q = P + \frac{1}{2}H = 0.40$
A_2A_2	60	$Q = 0.30$		
Total	200	$P + H + Q = 1$		$p + q = 1$

24 / 42

Outline

Introduction

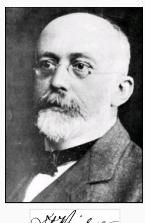
What is genetic variation?

Quantification of genetic variation

Hardy-Weinberg Equilibrium

25 / 42

Wilhelm Weinberg



- 1862-1937
- Born in Stuttgart, died in Tübingen
- Worked as Physician in Stuttgart
- >200 Publications on medical statistics and genetics
- His study was published in 1908
... *when Mendelian inheritance is viewed under the influence of panmixia.*

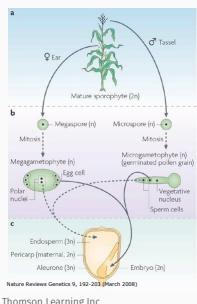
26 / 42

Hardy-Weinberg Equilibrium (HWE)

- Key assumption: Random mating or panmixia: Each individual has the same chance to mate with any other individual in the population
- Individuals do not fertilize themselves, i. e. they are allogamous
- The HWE allows to calculate the expected genotype frequencies from the allele frequencies.

27 / 42

The plant life cycle



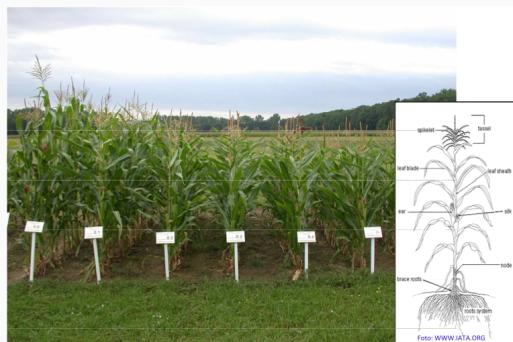
Nature Reviews Genetics 9, 192-203 (March 2008)

Thomson Learning Inc.

The gametophytes corresponds to the gametes that show random union in a population.

28 / 42

Example of an allogamous species: Maize



29 / 42

Random union of gametes

		Egg cells	
		A_1 (p)	A_2 (q)
Sperm cells	A_1 (p)	A_1A_1 (p^2)	A_1A_2 (pq)
	A_2 (q)	A_2A_1 (qp)	A_2A_2 (q^2)

30 / 42

Relationship between allele and genotype frequencies

$$\begin{array}{lll} \text{Genotype:} & A_1A_1 & A_1A_2 & A_2A_2 \\ \text{HW frequency:} & p^2 & 2pq & q^2 \end{array}$$

31 / 42

Temporal dynamics of HWE

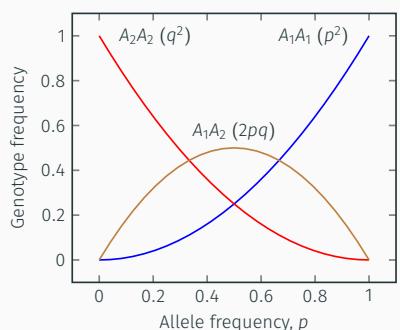
- **Father:** Genotype $A_1A_1 \rightarrow p = 1, q = 0$
- **Mother:** Genotype $A_2A_2 \rightarrow p = 0, q = 1$
- **Population:** $p = q = 0.5$

Generation	Genotypes			Allele frequencies
	A_1A_1	A_1A_2	A_2A_2	
0	0.5	0	0.5	$p = q = 0.5$
1	0	1	0	$p = q = 0.5$
2	0.25	0.5	0.25	$p = q = 0.5$

The HWE is achieved after two generations!

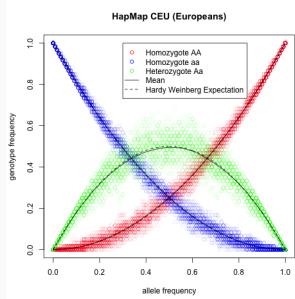
32 / 42

Relationship between allele and genotype frequencies



33 / 42

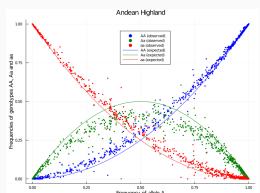
Distribution of genotype frequencies in humans



- 180 people from Utah with European ancestry
- 10,000 SNP markers; 3 dots per SNP

34 / 42

Distribution of genotype frequencies in maize landraces



- 222 accessions from the highlands of South America
- 982 SNP markers; 3 dots per SNP

35 / 42

Assumptions of HWE

The Hardy-Weinberg Equilibrium is modeled as a perfect probabilistic experiment ('Urn experiment').

Therefore:

- All individuals are diploid.
- There is only sexual recombination.
- Generations do not overlap.
- Allele frequencies are identical in both sexes.
- Mating is random → **Inbreeding**
- Population size is infinite → **Genetic drift**
- There is no mutation, migration and → **selection**.

All assumptions are required for the HWE!

36 / 42

Applications of HWE in plant breeding

- A **null hypothesis** for the expected number of homozygous and heterozygous genotypes
- Basis for quantitative genetics theory of expected phenotypic variation

37 / 42

Summary

- There are numerous types of genetic variation
- The Hardy-Weinberg equilibrium describes the relationship between genotype and allele frequencies
- The Hardy-Weinberg equilibrium makes numerous assumptions, some of which are biologically unrealistic.

38 / 42

Further reading

- [Becker, 2019], Chapter 6
- [Griffiths et al., 2020], Chapter 18
- [Acquaah, George, 2012], Chapter 7

39 / 42

Review Questions i

1. What is the importance of the analysis of genetic variation in plant breeding?
2. What are the key aspects of the central dogma of molecular biology?
3. What is the difference between a marker and a polymorphism?
4. What is the difference between an allele and a genotype?
5. What are the most important assumptions of the Hardy-Weinberg-Equilibrium?
6. Why does the Hardy-Weinberg-Equilibrium have so many assumptions?

40 / 42

Review Questions ii

7. Which assumption do you consider to be most unrealistic and what is going on in a population if it is violated?
8. Why does the HWE explain the persistence of recessive deleterious polymorphisms in a population?
9. What is the fate of dominant deleterious mutations in a population according to the HWE?

41 / 42

References i

- Acquaah, George (2012). *Principles of Plant Genetics and Breeding*. Wiley, 2nd edition.
- Becker, H. (2019). *Pflanzenzüchtung*. Verlag Eugen Ulmer Stuttgart, 3rd edition.
- Griffiths, A. J. F., Doebley, J., Peichel, C., and Wasserman, D. A. (2020). *Introduction to Genetic Analysis*. MacMillan International, 12th edition.
- Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., Li, C., Zhu, C., Lu, T., Zhang, Z., Li, M., Fan, D., Guo, Y., Wang, A., Wang, L., Deng, L., Li, W., Lu, Y., Weng, Q., Liu, K., Huang, T., Zhou, T., Jing, Y., Li, W., Lin, Z., Buckler, E. S., Qian, Q., Zhang, Q-F., Li, J., and Han, B. (2010). **Genome-wide association studies of 14 agronomic traits in rice landraces**. *Nature Genetics*, 42(11):961–967.

42 / 42



Population Genetics - Part 2

3501-211 Genetische Grundlagen der Pflanzenzüchtung

Prof. Karl Schmid
WS 2023/2024

Institute of Plant Breeding, Seed Science and Population Genetics
University of Hohenheim

1 / 45

Overview of lectures on population and quantitative genetics

Population Genetics

- The importance of genetic variation
- The Hardy-Weinberg Equilibrium
- Genetic drift
- Inbreeding
- Selection
- (Recombination and Linkage)

Quantitative Genetics

- Basic model of quantitative genetics
- Variance components and their estimation
- Phenotypic selection

2 / 45

Outline

- Genetic drift
- Inbreeding
- Gene frequency changes under selection

3 / 45

Summary: Population genetics - Part 1

- There are various types of genetic variation
- The Hardy-Weinberg equilibrium describes the relationship between genotype and allele frequencies
- Inbreeding increases the frequency of homozygous genotypes and causes inbreeding depression

$$\begin{array}{ccc} \text{Genotype freqs} & \xrightarrow{\quad p = P + \frac{1}{2}H, q = Q + \frac{1}{2}H \quad} & \text{Allele freqs} \\ P, H, Q & \xleftarrow{\quad \text{HWE: } P = p^2, H = 2pq, Q = q^2 \quad} & p, q \end{array}$$

4 / 45

Processes that change genotype or allele frequencies

- Genetic drift
- Inbreeding
- Selection
- (Population structure)
- (Migration or gene flow)
- (Mutation)

Recombination changes the combination of alleles.

5 / 45

The Hardy Weinberg Equilibrium (HWE) states that if all assumptions are fulfilled, allele frequencies and genotype frequencies in a population remain constant.

However, there are several processes that change allele frequencies, and for this reason contribute to the evolution of populations.

In addition, as we will learn in the class on quantitative genetics, the frequency of alleles at genes controlling quantitative traits also influences the level of phenotypic variation.

Therefore, changes in allele frequencies at genes lead to changes in phenotypic variation in a population, and it is necessary to study the processes that change allele and genotype frequencies.

Outline

Genetic drift

Inbreeding

Gene frequency changes under selection

6 / 45

Genetic drift

- Breeding populations are often small: $N < 100$
- Genetic drift causes the **loss of alleles** → reduced genetic variation
- The smaller the population, the stronger is genetic drift

7 / 45

Genetic drift is the process of random evolution.

It arises when the HWE assumption of an infinite population size is violated. Then, the sampling of a **finite** number of gametes (which carry one allele each) may lead to a different allele frequency in the next generation.

One consequence is that allele frequencies change between generations, in contrast to the prediction of HWE.

The direction of change of allele frequencies is random.

Genetic drift plays a substantial role in breeding population because they tend to be small. One important consequence is that genetic drift always causes a loss of alleles over time, and the rate of loss is higher with smaller populations.

What is the role of chance in reproduction?

Example:

- Population in HWE with alleles A_1 and A_2 ($p = q = 1/2$), genotype frequencies $\frac{1}{4}A_1A_1$, $\frac{1}{2}A_1A_2$ and $\frac{1}{4}A_2A_2$.
- Four individuals start a new population by randomly drawing eight gametes
- Probability that the new population consists of individuals of genotype A_1A_1 : $(\frac{1}{4})^4 = \frac{1}{256}$
- Probability that the new population is fixed for any one of two alleles: $2(\frac{1}{256}) = \frac{1}{128}$
- Change in allele frequencies is possible by drift: E.g. a population with genotypes $A_1A_1, A_1A_1, A_1A_1, A_1A_2$ (corresponds to $p = 7/8 = 0.875$ and $q = 1/8 = 0.125$)

8 / 45

The probability that the population is fixed for any one of two alleles is calculated by adding up the individual probabilities of fixation for one allele. Since a population can be fixed only for one of two alleles at a time, the two events (fixation for either A_1 or A_2) are not independent events, and therefore the total probability for the fixation of either one of two alleles is the sum of the individual probabilities (**addition rule** of probability theory).

The example shows how a finite population size leads to a change in allele frequency, and in this specific example, to a loss of genetic variation just by chance.

The simple numerical example nicely demonstrates that within a single generation the frequency of allele A_1 may change from 0.5 to 0.875.

The statistical theory underlying this sampling process is **binomial sampling**

Properties of genetic drift

- Genetic drift is **random** with respect to its direction.
- Average gene frequency of multiple populations with genetic drift:

$$\bar{q} = q_0$$

where q_0 is frequency of allele A_2 in the base population

- Variance of allele frequencies in different subpopulations after one generation

$$\sigma_q^2 = \frac{p_0 q_0}{2N}$$

and after t generations

$$\sigma_q^2 = p_0 q_0 \left[1 - \left(1 - \frac{1}{2N} \right)^t \right]$$

9 / 45

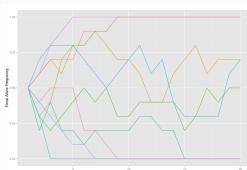
The stochastic nature of allele sampling makes drift random with respect to its direction.

If one would sample multiple individual populations with initial allele frequencies of q at time 0, the average allele frequencies would be \bar{q} . The variance of allele frequencies over the individual populations is the binomial sampling variance. Over time the value for the variance increases.

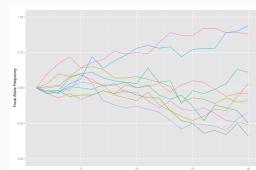
Genetic variation in small breeding populations

Monte-Carlo simulations of genetic drift with 10 subpopulations and two different population sizes

$N = 10$



$N = 100$



10 / 45

The two figures show a Monte Carlo simulation of genetic drift, which includes 10 independent populations in each simulation.

The change in allele frequencies over time is much stronger in the smaller than in the larger population. In addition, in the simulations with the smaller populations a substantial proportion of populations becomes fixed for either allele A_1 or A_2 .

Importance of genetic drift for plant breeding

- Genetic drift **always** causes the loss of genetic variation
- Small breeding populations experience strong genetic drift
- Theory of genetic drift can be used to calculate **minimal population size** in a breeding program to avoid genetic drift

The loss of variation is caused by the fixation of alleles.

11 / 45

Outline

Genetic drift

Inbreeding

Gene frequency changes under selection

12 / 45

Rare alleles occur mostly as heterozygotes

- Relative frequency of heterozygous genotypes with A_2 :
$$\frac{A_1A_2}{A_2A_2} : \frac{2pq}{q^2} \approx \frac{2p}{q}$$
- Example 1: $p = 0.999, q = 0.001$
Then,
$$A_1A_2/A_2A_2 = 1998 : 1$$
- Example 2: $p = q = 0.5$
Then,
$$A_1A_2/A_2A_2 = 2 : 1$$

Question:

What does this mean for rare, recessive, deleterious alleles?

One important consequence of HWS equilibrium is the existence of recessive deleterious (= disadvantageous) alleles in a population because if they are rare (i.e., their frequency is reduced by selection), they are mainly present in heterozygous genotypes (= carriers). As heterozygotes their deleterious effects are not expressed in the phenotype and therefore can not be "seen" and subsequently removed by selection.

13 / 45

Example calculation

- Assume: $p = 0.99, q = 0.01$
- Calculate: P, H, Q

Solution:

Genotype A_1A_1	$p^2 = 0.9801$
Genotype A_1A_2	$2pq = 0.0198$
Genotype A_2A_2	$q^2 = 0.0001$

14 / 45

The heterozygous genotype is called “carrier” of the recessive deleterious allele.

If this allele is rare, a much larger proportion of the allele is present in the heterozygous than the homozygous genotype and therefore not ‘visible’ to selection.

Why is inbreeding important?

- Rare, deleterious alleles mostly occur as **heterozygotes** (\rightarrow HWE)
- Self-fertilization increases **level of homozygosity**
- Deleterious homozygous mutations are expressed \rightarrow **Inbreeding depression**

15 / 45

The risk caused by inbreeding, which is also called inbreeding depression is higher in outcrossing crops such as maize and rye than in inbreeding crops such as wheat, rice, soybean and barley.

Expression of recessive deleterious mutations

Example: S_1 population of the Rheintaler Ribelmais maize landrace

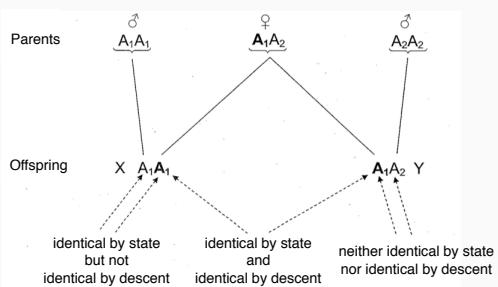


Photo: Benjamin Kogler, RhyTOP, Switzerland

16 / 45

In this population, which was produced by one generation of self-fertilization, a small proportion of plants do not produce functional chloroplasts, which are visible as white plants and their frequency in the population suggests they are recessive homozygous genotypes of a deleterious allele in an unknown gene.

Identity by descent (IBD)



17 / 45

The figure introduces the central concept of identity by descent in population genetics.

Identity of Alleles

Two definitions:

- **Identity by state** (IBS): Two alleles have the same state
- **Identity by descent** (IBD): Two alleles have the same common ancestor

Probability that two alleles in a population are IBD in the previous generation corresponds to **inbreeding coefficient**:

$$f = \Pr(\text{IBD}) = \frac{1}{2N}$$

- N : Number of individuals in a population
- 2N : Number of chromosomes in a population of diploid individuals

18 / 45

The inbreeding coefficient indicates the probability and *any* two randomly selected alleles from the same population originated from gametes produced from the same parental chromosome. It assumes a model of $2N$ parental chromosomes from which an infinite number of gametes is produced, and from them $2N$ gametes are randomly selected to generate the n diploid individuals in the next generation.

Therefore, the probability to obtain two IBD alleles is the probability to pick any single chromosome from a population ($P = 1$) multiplied by the probability that a second randomly drawn chromosome is IBD to the first one:

$$f = \Pr(\text{IBD}) = 1 \times \frac{1}{2N}$$

This calculation is an application of the multiplication rule of probability theory: The total probability of the probabilities of two independent events (first and second draw of a chromosome in an infinite gamete pool) is the product of the two individual probabilities.

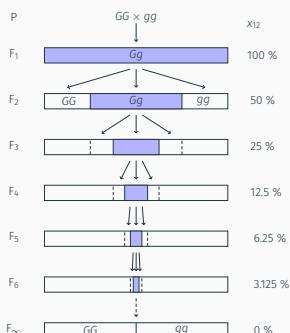
Loss of heterozygosity by inbreeding

Assumptions:

- Two homozygous parents, A₁A₁ and A₂A₂
- Population of heterozygous F₁ individuals, A₁A₂
- Further reproduction by recurrent self-fertilization: No outcrossing between individuals!
- Random union of gametes within individuals

19 / 45

Loss of heterozygosity by inbreeding



20 / 45

Loss of heterozygosity by inbreeding

After 5 generations of self-fertilization, 3.125% of individuals are heterozygous

Allele frequencies remain constant, in this case at $p = q = 0.5$!

Inbreeding reduces the frequency of heterozygous genotypes

Inbreeding makes recessive alleles homozygous

21 / 45

H is the fraction of heterozygous genotypes, as shown in blue in the previous figure.

Without inbreeding, the proportion of heterozygous genotypes in a population does not change.

If $F = 0$, there is no inbreeding and with $F = 1$ complete inbreeding.

Estimation of the inbreeding coefficient

- Inbreeding measured as decay of heterozygosity
- Definition of inbreeding coefficient F in generation t :

$$F_t = \frac{H_0 - H_t}{H_0}$$

- F ranges from 0 to 1
- Without inbreeding: $H_t = H_0$. Therefore, $F = 0$

22 / 45

Example:

Assume non-inbred parents in F_1 generation: $H_0 = 1$

In F_6 after 5 generations of self-fertilization:

- 3.125% of individuals are heterozygous at a given locus

$$F_t = \frac{H_0 - H_t}{H_0} = \frac{1 - 0.03125}{1} = 0.96875$$

23 / 45

Inbreeding depression - Why is inbreeding a problem?

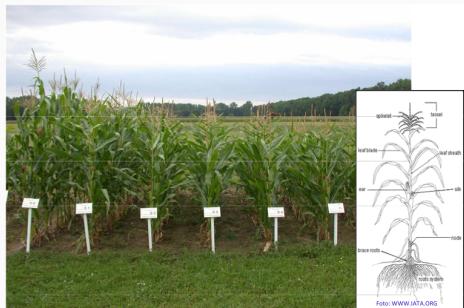
Several hypotheses exist, e.g., **dominance hypothesis**

- Expression of deleterious mutations: A_2 is recessive deleterious
- A_1A_2 Genotype: A_2 is not expressed
- A_2A_2 Genotype: A_2 is expressed → 'Disease'!
- Inbreeding increases the relative frequency of homozygous genotypes

24 / 45

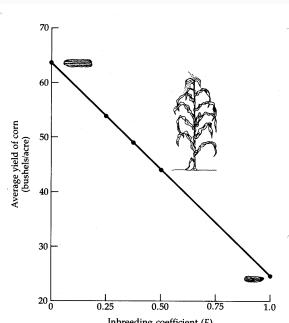
Example of an allogamous species: Maize

Effects of inbreeding are visible by plant height



25 / 45

Effect of inbreeding on yield in maize



Hartl and Clark, Principles of Population Genetics, 2007

26 / 45

Outline

Genetic drift

Inbreeding

Gene frequency changes under selection

27 / 45

Definition of selection

Natural selection: Selection of genotypes that are better adapted to the habitat

Artificial selection: Selection of genotypes that meet the breeding goal

The relative gametic contribution of a genotype to the next generation is the **reproductive success** or **fitness**

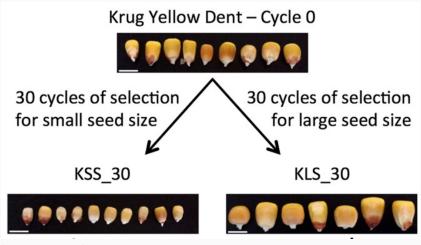
The frequency of **favorable** alleles increase, and the superior genotypes become more common.

⇒ Selection leads to **gene frequency changes** in both natural evolution and breeding.

28 / 45

Selection for smaller and larger maize kernels

- Recurrent selection of open pollinated populations for 30 cycles (= generations).
- Truncation selection: Only plants with seeds above (large seed) or below (small seed) are allowed to reproduce
- What happens with allele frequencies of genes influencing seed size?



[Hirsch et al., 2014]

One of the interesting questions in this experiment is which polymorphisms changed their allele frequencies in response to selection, and which genes they are located.

29 / 45

The basic model of selection

The following model is based on a **single** locus!

Genotype:	A_1A_1	A_1A_2	A_2A_2
Frequency in newborn:	p^2	$2pq$	q^2
Survival probability:	w_{11}	w_{12}	w_{22}
Product	p^2w_{11}	$2pqw_{12}$	q^2w_{22}
Frequency after selection:	$\frac{p^2w_{11}}{\bar{w}}$	$\frac{2pqw_{12}}{\bar{w}}$	$\frac{q^2w_{22}}{\bar{w}}$

Mean population fitness:

$$\bar{w} = p^2w_{11} + 2pqw_{12} + q^2w_{22}$$

A correction of the individual genotype frequencies *after selection* by dividing with the mean population fitness is required because only then the individual genotype frequencies add up to 1.

30 / 45

Three types of selection

The allele frequency change before and after selection depends on the fitness coefficients:

- Directed selection: $w_{11} > w_{12} > w_{22}$ or $w_{11} < w_{12} < w_{22}$
- Balanced selection: $w_{11} < w_{12} > w_{22}$
- Disruptive selection: $w_{11} > w_{12} < w_{22}$

Possible fates of alleles:

- Fixation of A_1 in the population ($p \rightarrow 1$) (directed selection)
- Loss of A_1 from the population ($p \rightarrow 0$) (directed selection)
- Equilibrium at a certain allele frequency ($p \rightarrow \hat{p}$) (balanced selection)
- No change (in equilibria)

31 / 45

Relative Fitness

Genotype:	A_1A_1	A_1A_2	A_2A_2
Viability:	w_{11}	w_{12}	w_{22}
Relative viability:	1	w_{12}/w_{11}	w_{22}/w_{11}
Relative Fitness:	1	$1 - hs$	$1 - s$

with $1 - hs = w_{12}/w_{11}$ and $1 - s = w_{22}/w_{11}$.

- s : Selection coefficient - the fitness difference between the two homozygous genotypes
- h : Heterozygote effect

32 / 45

The effect of heterozygotes

$h = 0$	A_1 dominant, A_2 recessive
$h = 1$	A_1 recessive, A_2 dominant
$0 < h < 1$	Incomplete dominance: $w_{11} > w_{12} > w_{22}$
$h < 0$	Overdominance: $w_{11} < w_{12} > w_{22}$
$h > 1$	Underdominance $w_{11} > w_{12} < w_{22}$

- Change of allele frequencies:

$$\Delta_s p = \frac{pq[s(ph + q(1-h))]}{\bar{w}}$$

- Mean fitness:

$$\bar{w} = 1 - 2pqhs - q^2s$$

33 / 45

A numerical example

Assumption: A cross-fertilized population segregating for a monogenic dominant disease resistance:

Allele:	R	resistant	r	susceptible
Frequency:	p	0.6	q	0.4

Selection coefficient: $s = 0.3$

The following slide shows changes in allele and genotype frequencies after one generation of selection...

34 / 45

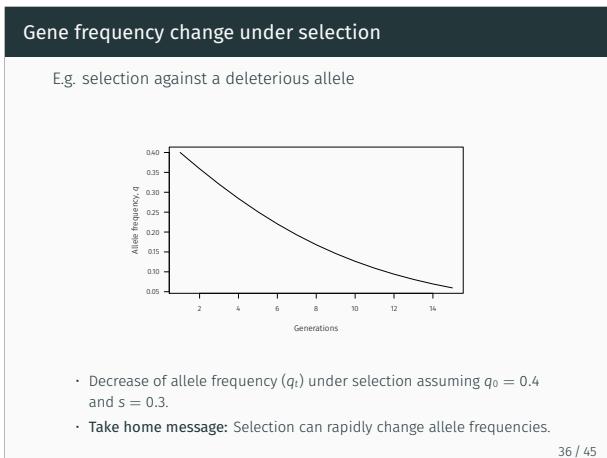
Genotype	<i>RR</i>	<i>Rr</i>	<i>rr</i>	Total
Phenotype	resistant		susceptible	
Freq. before infection	p^2 0.36	$2pq$ 0.48	q^2 0.16	1
Fitness (affected by disease)	1 1	1 1	1 - s 0.7	-
Proportion after the epidemic	p^2 0.36	$2pq/\bar{w}$ 0.48	$(1-s)q^2/\bar{w}$ 0.112	\bar{w}
Frequency in next generation before infection	p'^2/\bar{w} 0.378	$2p'q'/\bar{w}$ 0.504	$(1-s)q'^2/\bar{w}$ 0.118	1

Population fitness: $\bar{w} = p^2 + 2pq + (1-s)q^2 = 1 - sq^2 = \mathbf{0.952}$
New allele frequency: $q' = [pq + 1(1-s)q^2]/\bar{w} = \mathbf{0.370}$
Change in allele frequency: $\Delta q = q' - q = -\frac{sq^2}{1-sq^2} = \mathbf{-0.030}$

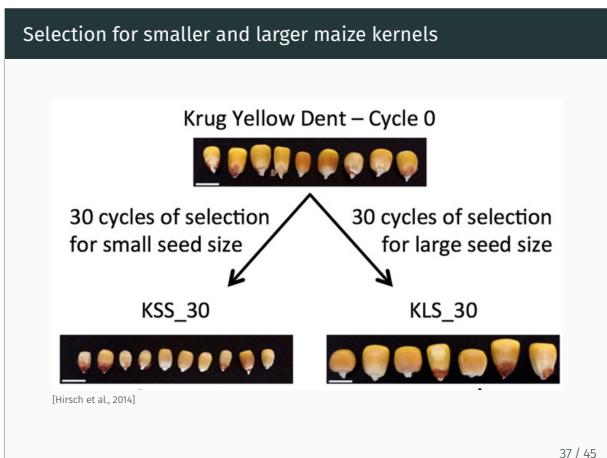
Genotype frequency in next generation before infection

$(p')^2$ 0.397	$2p'q'$ 0.466	$(q')^2$ 0.137	1
--------------------------	-------------------------	--------------------------	---

35 / 45



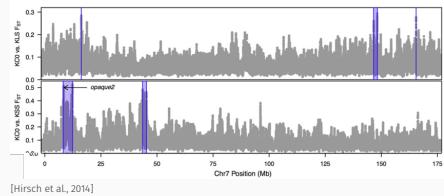
36 / 45



37 / 45

Application in plant breeding

- Differences in SNP allele frequencies between original (K_{C_0}) and selected populations (K_{SS} and K_{LS})
- F_{ST} measures allele frequencies between populations ($F_{ST} = 0$: No differences, $F_{ST} = 1$: Different alleles fixed)



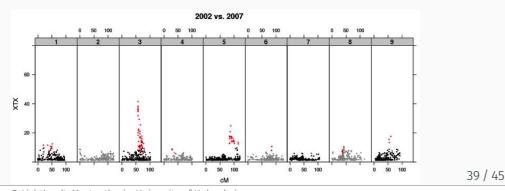
Violet regions: Statistical support of selection-driven allele frequency change (99.9% outlier).

38 / 45

SNP: Single nucleotide polymorphism. A polymorphism at a single nucleotide position in a genome.

Application in plant breeding

- Introgression** of new disease resistance genes from wild relatives
- Disease: Rhizomania in sugar beet, a virus disease
- Recurrent selection:** Cross in each generation and select resistant genotypes
- Differences in allele frequencies per SNP marker after several rounds of selection:



C. Lichthardt, Master thesis, University of Hohenheim

39 / 45

X^2X is a measure of genetic differentiation, similar to the F_{ST} measure above.

Identifying regions with strong genetic differentiation in response to selection allows the identification of genes, which played a role in selection.

Selection in plant breeding

- Selection of favorable traits is usually very strong
- Selection on genetic variation of causal genes for a trait is also very strong
- In the long term, selection leads to a loss of genetic variation (fixation of alleles)

40 / 45

Summary

- Genetic drift is the random change in allele frequencies
- Inbreeding increases the frequency of homozygous genotypes and causes inbreeding depression
- Inbreeding increases the proportion of homozygous genotypes
- There are three types of selection: directed, balanced and disruptive
- Selection is most efficient with intermediate allele frequencies.
- Selection on rare alleles is inefficient (if they are recessive).

41 / 45

Further reading

- [Becker, 2019], Chapter 6
- [Griffiths et al., 2020], Chapter 18
- [Acquaah, George, 2012], Chapter 7

42 / 45

Review Questions i

1. Does a change in allele frequencies also change genotype frequencies in a population?
2. What is the main cause of genetic drift and which factors determines its level?
3. What are the key properties of genetic drift?
4. Do you think genetic drift is a problem in breeding populations?
5. What is the difference of alleles that are identical by their state and identical by their descent?
6. How is the inbreeding coefficient defined?
7. Does inbreeding change allele frequencies, genotype frequencies, or both in a population?

43 / 45

Review Questions ii

8. What is the effect of inbreeding on the expression of deleterious alleles?
9. What are the different types of selection and what is their effect on allele frequencies?
10. What is the effect of heterozygous genotypes on the result of selection?
11. Under which circumstances is selection particularly effective?

44 / 45

References i

- Acquaah, George (2012). *Principles of Plant Genetics and Breeding*. Wiley, 2nd edition.
- Becker, H. (2019). *Pflanzenzüchtung*. Verlag Eugen Ulmer Stuttgart, 3rd edition.
- Griffiths, A. J. F., Doebley, J., Peichel, C., and Wasserman, D. A. (2020). *Introduction to Genetic Analysis*. MacMillan International, 12th edition.
- Hirsch, C. N., Flint-Garcia, S. A., Beissinger, T. M., Eichten, S. R., Deshpande, S., Barry, K., McMullen, M. D., Holland, J. B., Buckler, E. S., Springer, N., Buell, C. R., de Leon, N., and Kaeplke, S. M. (2014). Insights into the Effects of Long-Term Artificial Selection on Seed Size in Maize. *Genetics*, 198(1):409–421.

45 / 45



Quantitative Genetics

3501-211 Genetische Grundlagen der Pflanzenzüchtung

Prof. Karl Schmid
WS 2023/2024

Institute of Plant Breeding, Seed Science and Population Genetics
University of Hohenheim

1 / 55

Overview of Lectures

Population Genetics

- The nature of genetic variation
- The Hardy-Weinberg Equilibrium
- Inbreeding
- Genetic drift
- Selection
- Recombination and Linkage

Quantitative Genetics

- Basic model of quantitative genetics
- Variance components and their estimation
- Phenotypic selection

2 / 55

Recapitulation

- From genotype to allele frequencies and back
- Importance of the Hardy-Weinberg Equilibrium
- Changes in genetic variation due to genetic drift and selection
- Effects of inbreeding and recombination

3 / 55

Learning goals

- Understand the nature of quantitative traits
- Effect of gene action on phenotypic variation
- How to model genetic and phenotypic variation
- How to determine and utilize the breeding value and variance for plant breeding

4 / 55

In this section, we aim to cover several critical learning goals. First, we will delve into understanding the nature of quantitative traits and how they manifest within populations. Following this, we will explore the effect of gene action on phenotypic variation, which is a key concept in the study of genetics. Additionally, we will look at methodologies for modeling both genetic and phenotypic variation, which are essential for predicting outcomes in breeding programs. Lastly, we will learn how to determine and effectively utilize the breeding value and variance, specifically within the context of plant breeding, to enhance our breeding strategies and achieve desired traits in crops.

Outline

Phenotypic traits

Polygenic inheritance

Gene action

Variance components of phenotypic traits

Heritability

Selection

5 / 55

Phenotypic traits

Polygenic inheritance

Gene action

Variance components of phenotypic traits

Heritability

Selection

6 / 55

"In this section, we aim to cover several critical learning goals. First, we will delve into understanding the nature of quantitative traits and how they manifest within populations. Following this, we will explore the effect of gene action on phenotypic variation, which is a key concept in the study of genetics. Additionally, we will look at methodologies for modeling both genetic and phenotypic variation, which are essential for predicting outcomes in breeding programs. Lastly, we will learn how to determine and effectively utilize the breeding value and variance, specifically within the context of plant breeding, to enhance our breeding strategies and achieve desired traits in crops."

7 / 55

Qualitative and quantitative traits in plants

Qualitative traits	Quantitative traits
Colour	Yield
Dwarfism	Size
Self-incompatibility	Growth
Amino acid composition	Fertility
Specific resistances	Protein (or fat content)
	Stress tolerance
	1000 kernel weight
	Concentration of metabolites, micronutrients, etc.
	Resistance against different pathogens

7 / 55

When discussing the characteristics of plants, it is crucial to distinguish between qualitative and quantitative traits, as each has a significant impact on plant breeding and cultivation.

Qualitative traits are typically those that can be classified into distinct categories based on their appearance or presence/absence of a particular attribute. For example, the color of a plant, whether it exhibits dwarfism, its ability to self-pollinate (self-incompatibility), the composition of amino acids within its structure, and specific resistances to certain environmental factors or pathogens are all qualitative traits. These traits are often controlled by a single gene or a small number of genes, which makes them relatively straightforward to identify and select for in breeding programs.

On the other hand, quantitative traits are those that exist along a continuum and cannot be easily classified into categories because they are influenced by multiple genes and often affected by environmental conditions. These include the yield of the plant, its overall size, rate of growth, fertility levels, and the content of proteins or fats it contains. Additionally, traits such as stress tolerance, the weight of a thousand kernels (which is a common measure in grain crops), and the concentration of metabolites or micronutrients also fall into this category. Furthermore, resistance to various pathogens can be a complex trait influenced by many genetic and environmental factors, making it quantitative.

In the context of plant breeding, these distinctions are crucial. Qualitative traits may be easier to select for due to their simplicity, but quantitative traits often have a more significant impact on the performance and viability of a crop. As such, advanced statistical and genetic tools are used to analyze these complex traits to improve crop varieties and to ensure that they can meet the demands of consumption and survival in diverse environments. Understanding these differences helps breeders create better strategies for the development of new plant varieties.

Example of a qualitative trait

Trait:	Glume color of barley	
Phenotype:		
Genotype:	bb	BB and Bb
Gene:	Glume color: B	
Alleles:	B : black, b : non-black	

8 / 55

Continuously varying traits: Body size



Means and standard deviations

- Males: 70.1 ± 3.0
- Females: 64.8 ± 2.7
- Combined: 67.6 ± 4.0

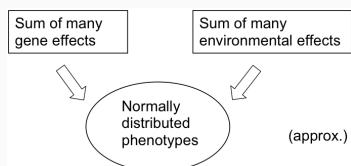
9 / 55

Types of genetic and phenotypic variation

Type of genetic variation	Environment	
	Homogenous	Heterogeneous
Uniform	Qualitative traits	
	No variation	
Monogenic variation		
Uniform	Quantitative traits	
	No variation	
Polygenic variation		

10 / 55

Origin of quantitative variation



Many quantitative traits are **normally distributed**.

Parameters of a normal distribution

- | | |
|-------------------|--|
| x | random variable (e.g., a vector of trait values) |
| μ | population mean = $E\{x\}$ |
| σ | standard deviation |
| $\sigma^2 = V(x)$ | Variance = $E\{(x - \mu)^2\}$ |

11 / 55

Properties of quantitative traits

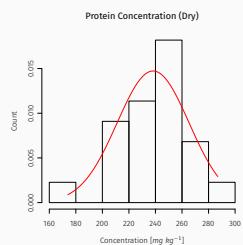
- (Many) trait values follow a **normal distribution**
- They are controlled by **multiple genes**
- The traits are affected by **environmental variation** as well
- The genes controlling traits are usually **unknown**
- Different quantitative traits can be **correlated**

12 / 55

Micronutrient and protein content in wild emmer wheat



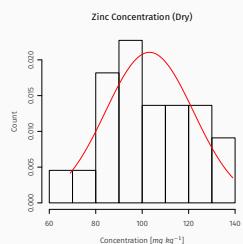
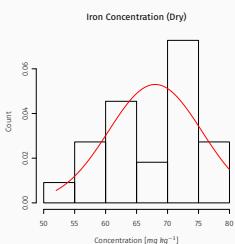
Foto: Zvi Peleg



Data source: [Peleg et al., 2008]

13 / 55

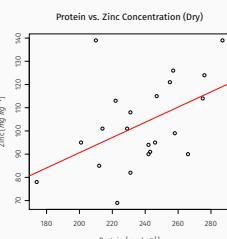
Micronutrient and protein content in wild emmer wheat



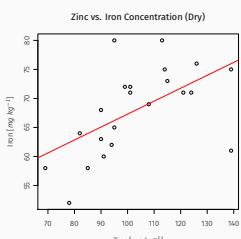
Data source: [Peleg et al., 2008]

14 / 55

Correlation between quantitative traits



$R^2 = 0.22, p = 0.028$



$R^2 = 0.32, p = 0.007$

15 / 55

Phenotypic traits

Polygenic inheritance

Gene action

Variance components of phenotypic traits

Heritability

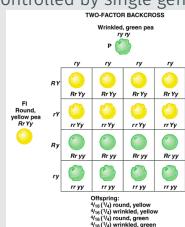
Selection

16 / 55

Mendelian versus quantitative genetics

Mendelian genetics

Simple, discrete traits controlled by single genes



Quantitative genetics

Complex traits controlled by multiple genes



How are Mendelian and quantitative genetics related?

17 / 55

Basic model of quantitative genetics: One locus

- Assumption: A single locus with 2 alleles, A and a
- Allele frequencies: $p = q = 0.5$
- Gametes: A and a
- Each A adds up one **phenotypic unit**, each a zero.
- Combinations in offspring:

	A	a
A	AA	Aa
a	aA	aa

- Three different genotypes: AA , Aa , aa
- Phenotypes: Add up **phenotypic units** (capital letters):

Phenotype class:	0	1	2
Frequency:	1/4	2/4	1/4

Phenotypes are in Hardy-Weinberg equilibrium frequencies!

18 / 55

Basic model of quantitative genetics: Two loci

- Two loci A and B control a trait with allele frequencies

$$A = a = B = b = 0.5$$

- Gamete combinations: AB , Ab , aB , ab
- 16 allele combinations in offspring:

	AB	Ab	aB	ab
AB	AABB	AABb	AaBB	AaBb
Ab	AABb	AAbb	AaBb	Aabb
aB	AaBB	AaBb	aaBB	aaBb
ab	AaBb	Aabb	aaBb	aabb

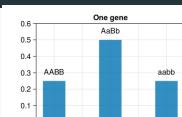
- Add up **phenotypic units** (capital letters):

Phenotype class:	0	1	2	3	4
------------------	---	---	---	---	---

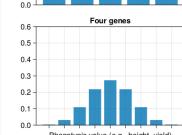
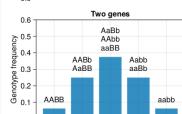
Frequency:	1/16	4/16	6/16	4/16	1/16
------------	------	------	------	------	------

19 / 55

Phenotypic values of traits

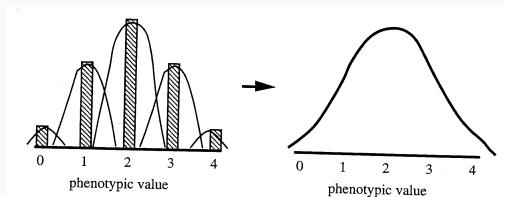


One gene
Variation in phenotypic values results from **additive genetic effects**: The phenotypic value is the sum of the effects of the individual alleles.



20 / 55

Effect of environment on trait variation



The environment introduces additional variation on the phenotype.

The phenotype distribution becomes similar to a normal distribution.

21 / 55

Key decisions in breeding

In plant breeding, an important goal is to improve a particular phenotypic (usually quantitative trait). In order to achieve this, the following questions need to be addressed:

1. To what degree is the trait inherited or determined by the environment?
2. How much variation in my breeding population is genetic?
3. What is the nature of the genetic variation? Additive, dominance, epistatic variation
4. How is the genetic variation organised? Refers to the number and effect of different genes on a trait

To answer these questions, we use a metric genetic model describing phenotypic variation.

22 / 55

Phenotypic traits

Polygenic inheritance

Gene action

Variance components of phenotypic traits

Heritability

Selection

23 / 55

A simple genetic model for describing phenotypic variation

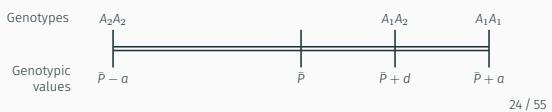
Goal: Estimate genetic effects (additive, dominance) by analyzing phenotypic variation

Assumptions:

- One locus, two alleles
- \bar{P} : Midparent value

Only two model parameters are necessary:

- a : half of phenotypic distance between A_1A_1 and A_2A_2
- d : Deviation of heterozygote from mean between homozygotes



24 / 55

Additive effect, a



Measures the deviation of the homozygous genotypes from the midparent value

Additive effect remains the same regardless of the allele with which it is combined

25 / 55

Dominance effect, d

Dominance refers to any degree of allelic interaction ranging from no dominance to complete over-dominance.

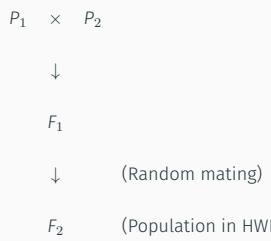


- $d = 0$: No dominance → Intermediate inheritance
- $d = a$: Complete dominance → Dominant-recessive inheritance
- $d > 0$ or $d < a$: Incomplete dominance
- $d > a$: Overdominance → Heterozygote has stronger phenotype than homozygotes
- $d < 0$: Negative dominance (recessiveness)

26 / 55

The genetic value of a population

Assume the following mating scheme:



- **Question:** What is the average phenotypic value of the F_2 Generation?
- **Assumption:** We estimated parameters a and d from the parental (P_1, P_2) and F_1 generations

27 / 55

Estimates of population mean of a trait, μ

- Assumption: Locus is in HWE
- Calculate **mean population value** of F_2 generation by multiplying each genotypic value with its HWE frequency:

$$\begin{aligned}\mu &= p^2(\bar{P} + a) + 2pq(\bar{P} + d) + q^2(\bar{P} - a) \\ &= \bar{P} + a(p - q) + 2pqd\end{aligned}$$

(Simplification by using the binomial formula and $p + q = 1$)

- Contribution of homozygote genotypes: $a(p - q)$
- Contribution of heterozygote genotypes: $2pqd$

28 / 55

Average effects of alleles

Problem: Parents transmit alleles, not genotypes

Genotypes in next generation result from transmitted alleles

Parameters $a, d, -a$ are functions of genotypes, not alleles

Goal: Estimate the **average effect of an allele, α** on population mean

29 / 55

Calculate average effect of allele A_1

- Crossing of a A_1A_1 individual with a population in HWE
- Probability that A_1 unites with A_1 is p
- Resulting A_1A_1 genotype has genotypic value $\bar{P} + a$
- Probability that A_1 unites with A_2 is q
- Resulting A_1A_2 genotype has genotypic value $\bar{P} + d$
- No A_2A_2 genotype is produced!
- Mean genotypic value of offspring that inherits allele A_1 is:

$$p(\bar{P} + a) + q(\bar{P} + d) = \bar{P} + pa + qd$$

- Average effect of A_1 is the deviation of offspring mean from population mean, μ :

$$\begin{aligned}\alpha_1 &= (\bar{P} + pa + qd) - \mu & (1) \\ &= (\bar{P} + pa + qd) - \bar{P} + a(p - q) + 2pqd \\ &= q[a + d(q - p)]\end{aligned}$$

30 / 55

Average effect of alleles

- Allele A_1 : $\alpha_1 = q[a + d(q - p)]$
- Allele A_2 : $\alpha_2 = -p[a + d(q - p)]$
- The effect of an allele is the effect on the trait mean of the individuals that inherit the allele
- Definition is relevant to selection in populations
- Selection aims to change population mean
- Change in the population mean when selection favors one allele over another → Direct function of the average effects of alleles

31 / 55

Average effect of individuals

- Add the effects of the two alleles:

$$\begin{aligned}A_1A_1 &= 2\alpha_1 \\ A_1A_2 &= \alpha_1 + \alpha_2 \\ A_2A_2 &= 2\alpha_2\end{aligned}$$

- These values are called **breeding values**
- They describe how genotypes change the phenotypic values in a population to their genotypic value given by a and $-a$

32 / 55

Phenotypic traits
Polygenic inheritance
Gene action
Variance components of phenotypic traits
Heritability
Selection

33 / 55

Components of phenotypic value

The phenotypic value of an individual, P , is therefore the sum of genetic value (G) and environmentally determined value (E):

$$P = G + E$$

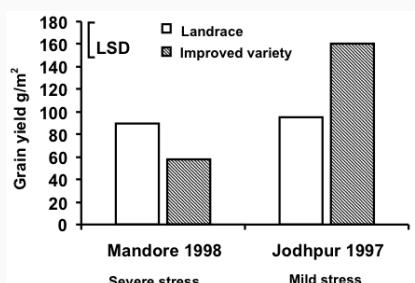
Many genes show approximately normally distributed traits!

Genotype-by-environment (GxE) interactions contribute to phenotypic value

34 / 55

Example of GxE interaction

Two varieties of pearl millet cultivated at two locations



35 / 55

Variance of a phenotypic trait

The total variance of a trait is the sum of its **variance components**:

$$V_P = V_G + V_E + V_{GE}$$

with

- V_P : Total phenotypic variance
- V_G : Genetic variance
- V_E : Environmental variance
- V_{GE} : Variance associated with GxE interactions

36 / 55

Variance components of genetic variation

Genetic variance can be split into different variance components due to different gene actions:

$$V_G = V_A + V_D + V_I$$

- V_A : Additive variance
- V_D : Dominance variance
- V_I : Interaction variance

V_A is the main cause of resemblance between relatives and primary determinant of the response to selection.

Total phenotypic variance:

$$V_P = V_A + V_D + V_I + V_E + V_{GE}$$

37 / 55

Variance components can be estimated from specially designed experiments

Phenotypic traits

Polygenic inheritance

Gene action

Variance components of phenotypic traits

Heritability

Selection

38 / 55

Properties of heritability

Heritability: The proportion of genetic variation compared to the total phenotypic variation

- **Narrow sense heritability:** $h^2 = V_A/V_P$
- **Broad sense heritability:** $H^2 = V_G/V_P$
- The heritability coefficient is specific for a
 - Population
 - Trait
 - Range of environments
- Heritability coefficients of complex traits are lower than those of simpler traits (see next slide)

39 / 55

Estimates of heritability

Species	Trait	h^2	H^2
Maize	Grain yield	0.19	0.32
	Plant height	0.57	0.67
Pig	Litter size	0.05	-
	Weight gain	0.40	-

40 / 55

Methods for estimating heritability

Two main types of methods for estimating heritability:

- Variance component method
- Parent offspring regression

Variance component method

1. Estimate individual variance components in field trials:
 V_P, V_A, V_E
2. Calculate $h^2 = V_A/V_P$ or $H^2 = V_G/V_P$

Parent-offspring regression

1. Determine phenotype of many pairs of parents and offspring
2. Calculate a regression coefficient of parent midvalues and offspring phenotype values: $h^2 = \beta_{OP} = V_A/V_P$

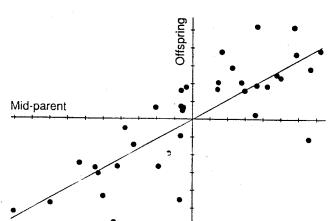
41 / 55

Regression of wing lengths in Drosophila



Mid-parent method: Take average of parent values:

$$h^2 = \beta_{OP} = 0.58$$



42 / 55

Phenotypic traits

Polygenic inheritance

Gene action

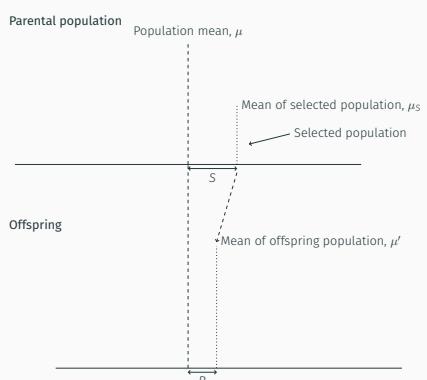
Variance components of phenotypic traits

Heritability

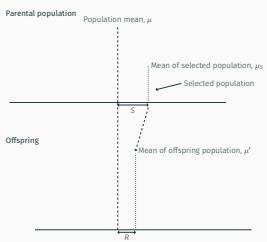
Selection

43 / 55

Model of truncation selection



44 / 55



μ = Mean value of parental population

μ_s = Mean value of group selected for reproduction

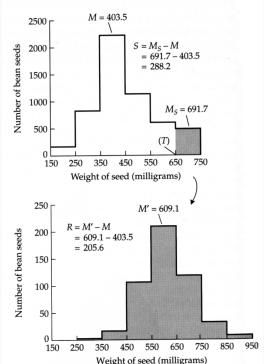
μ' = Mean value of offspring,

$$\text{Selection differential: } S = \mu_s - \mu$$

$$\text{Response to selection: } R = \mu' - \mu$$

45 / 55

Truncation selection in *Phaseolus* beans



46 / 55

Relationship between selection and heritability

- Recall that heritability can be seen as regression of offspring on mid-parent trait value
- Selection differential between generations:

$$\begin{aligned}\mu' &= \mu + h^2(\mu_s - \mu) \\ \mu' - \mu &= h^2(\mu_s - \mu)\end{aligned}$$

- Response to selection:

$$R = h^2 S$$

- Realised heritability:

$$h^2 = R/S$$

- Phaseolus example (seed weight):

$$h^2 = R/S = 205.6/288.2 = 0.713$$

47 / 55

What determines selection gain?

Three parameters are important:

- level of genetic variation in the population
- reliability by which this variation can be recognized
- strength of selection

The breeding equation uses these parameters to describe the response to selection:

$$R = ih\sigma_g$$

- i : Selection intensity (proportion of individuals selected)
- h : Square root of heritability, h^2
- σ_g : The square root of genetic variance

This equation is called the **Breeder's Equation**

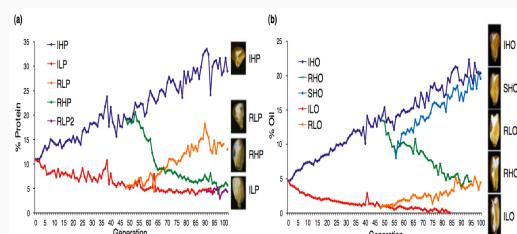
48 / 55

The Illinois Long-Term Selection Experiment in Maize

- Start in 1896 with 163 ears from Burr's White variety
- 24 highest scoring and 12 lowest scoring for selection
- Selection for high/low oil and seed protein
- Normal values for oil: 4-6% and protein: 8-12%
- Genetic variation still present after 100 generations!
- Selection responses compared to starting population:
 - Positive direction: > 20 standard deviations
 - Negative direction: 4 standard deviations
- Such extreme changes are not possible with mutagenesis or transgenic experiments

49 / 55

The Illinois Long-Term Selection Experiment in Maize



[Moose et al., 2004]

50 / 55

Conclusion - Quantitative Genetics

- Phenotypic variation is determined by genetic and environmental variation
- Genetic variation is partitioned into additive, dominance and interaction variation
- A genetic model allows to estimate genetic parameters from phenotypic variation
- The breeding value of individuals is a useful concept to estimate the effect of genetic variation on a population
- Variance components and heritability are useful concepts to describe the genetic composition of a population
- The breeder's equation describes the change of the phenotypic mean of a population in response to selection

51 / 55

Further reading

- [Becker, 2019], Chapter 7 and 8

52 / 55

Review questions i

1. Which different types of traits exist? Which ones do you think are most frequent and why do you expect this?
2. What is the link between Mendelian and quantitative genetics?
3. Which aspects of the underlying genes and genetic architecture of a trait does the genetic model of phenotypic variation describe? Which not?
4. Why is additive variance the most important type of variance for plant breeding?
5. What are variance components and how are they defined?
6. Why is a knowledge of variance components useful in plant breeding?

53 / 55

Review questions ii

7. What are the implications if a phenotypic trait has a low heritability?
8. Why is the heritability coefficient specific for a population, for a trait, and for different environments?
9. What is the selection differential?
10. Why is the knowledge of heritability important and useful for estimating the effects of selection?

54 / 55

References i

-  Becker, H. (2019). *Pflanzenzüchtung*. Verlag Eugen Ulmer Stuttgart, 3rd edition.
-  Moose, S. P., Dudley, J. W., and Rocheford, T. R. (2004). **Maize selection passes the century mark: a unique resource for 21st century genomics.** *Trends in Plant Science*, 9(7):358–364.
-  Peleg, Z., Saranga, Y., Yazici, A., Fahima, T., Ozturk, L., and Cakmak, I. (2008). **Grain zinc, iron and protein concentrations and zinc-efficiency in wild emmer wheat under contrasting irrigation regimes.** *Plant and Soil*, 306(1):57–67.

55 / 55