Crop Genetics

Walter Suza (Editor); Kendall Lamkey (Editor); William Beavis; Arden Campbell; Jode Edwards; Shui-Zhang Fei; Thomas Lübberstedt; Laura Merrick; and Deborah Muenchrath

Gretchen Anderson; Todd Hartnell; Andy Rohrback; Tyler Price; Glenn Wiedenhoeft; and Abbey Elder

Iowa State University Digital Press
Ames, Iowa

How to cite this publication:

Suza, W., & Lamkey, K. (Eds.). (2023). Crop Genetics. Iowa State University Digital Press. DOI: 10.31274/isudp.2023.130

This is a publication of the
Iowa State University Digital Press
701 Morrill Rd, Ames, IA 50011
https://www.iatatedigitalpress.com
digipress@iastate.edu

CONTENTS

About the PBEA Series

Background

The Plant Breeding E-Learning in Africa (PBEA) e-modules were originally developed as part of the Bill & Melinda Gates Foundation Contract No. 24576.

Building on Iowa State University's expertise with online plant breeding education, the PBEA e-modules were developed for use in curricula to train African students in the management of crop breeding programs for public, local, and international organizations. Collaborating with faculty at Makerere University in Uganda, University of KwaZulu-Natal in South Africa, and Kwame Nkrumah University of Science and Technology in Ghana, our team created several e-modules that hone essential capabilities with real-world challenges of cultivar development in Africa using Applied Learning Activities. Our collaboration embraces shared goals, sharing knowledge and building consensus. The pedagogical emphasis on application produces a coursework-intensive MSc program for Africa.

The authors of this textbook series adapted and built upon the PBEA modules to develop a series of textbooks covering individual topic areas. It is our hope that this project will facilitate wider dissemination and reuse of the PBEA modules' content.

Explore the Series

- Crop Genetics
- Molecular Plant Breeding (coming soon!)
- Crop Improvement (coming soon!)
- Quantitative Methods (coming 2023)
- Additional books to be announced later

Reproduction in Crop Plants

Deborah Muenchrath; Arden Campbell; Laura Merrick; Thomas Lübberstedt; and Shui-Zhang Fei

Introduction

Most agronomic and horticultural crop species are **angiosperms**. Angiosperms are vascular plants that produce their seeds enclosed in a matured ovary, a fruit; the fruit arises from a flower. In contrast, some tree crops—such as pine or spruce—are **gymnosperms**, which are vascular plants possessing "naked seeds" that are not enclosed within fruit structures. There are two groups of angiosperms: **monocots** and **dicots**. Although all angiosperms have some reproductive features in common, species vary in their mode of reproduction. A species' reproductive mode is fundamental to the methods applied to develop improved cultivars.

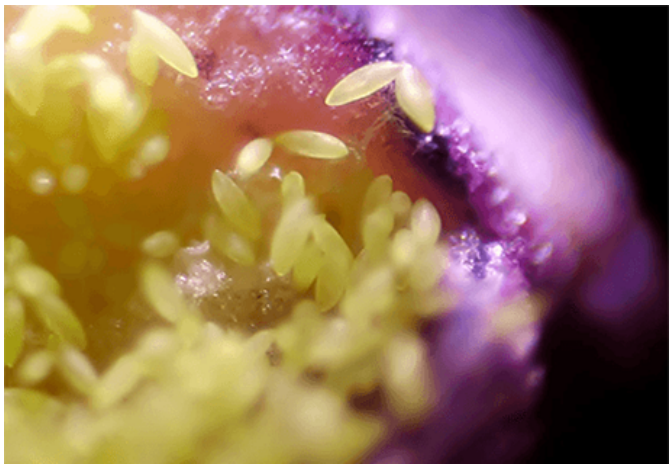


Fig. 1 Pollen grains on a Hippeastrum flower. Photo by Krucku, licensed under CC BY-SA 3.0 via Wikimedia Commons.

Learning Objectives

- Review hereditary mechanisms and flower anatomy.
- Understand the sexual reproduction processes of pollination, fertilization, and seed development.
- Become familiar with asexual reproduction.
- Comprehend the implications of reproductive mode for crop breeding strategies.

Hereditary Mechanisms

Heredity, Genotypes, and Phenotypes

Plant breeders take advantage of the mechanisms of **heredity** to develop and maintain cultivars. The observable characteristics and performance of cultivar (of a plant), its **phenotype**, is the result of the cultivar's **genotype**, as influenced by the environment. In other words, phenotype is a function of both genotype and environment, plus the interaction between genotype and environment.

$$P = G + E + (G \times E), \text{ where:}$$

$$P = \text{Phenotype};$$
$$G = \text{Genotype};$$
$$E = \text{Environment}$$

Fundamental to effective and efficient plant breeding is an understanding of the hereditary mechanisms that affect genotype:

- Nuclear division and chromosomes
- Modes of reproduction

Deoxyribonucleic Acid (DNA) And Chromosomes

Every cell nucleus contains the genetic material of the cell, **deoxyribonucleic acid** (**DNA**), located in **chromosomes**. Each chromosome is a single DNA molecule. Associated with the DNA are special proteins called histones (see the round yellow shapes in Fig. 2 around which the DNA is wound like "beads on a string" along the chromosome) that are related to the organization of the DNA, as well as enzymes involved in replication of the DNA strand. In plant cells, DNA and associated genetic information are mainly located in nuclear chromosomes.

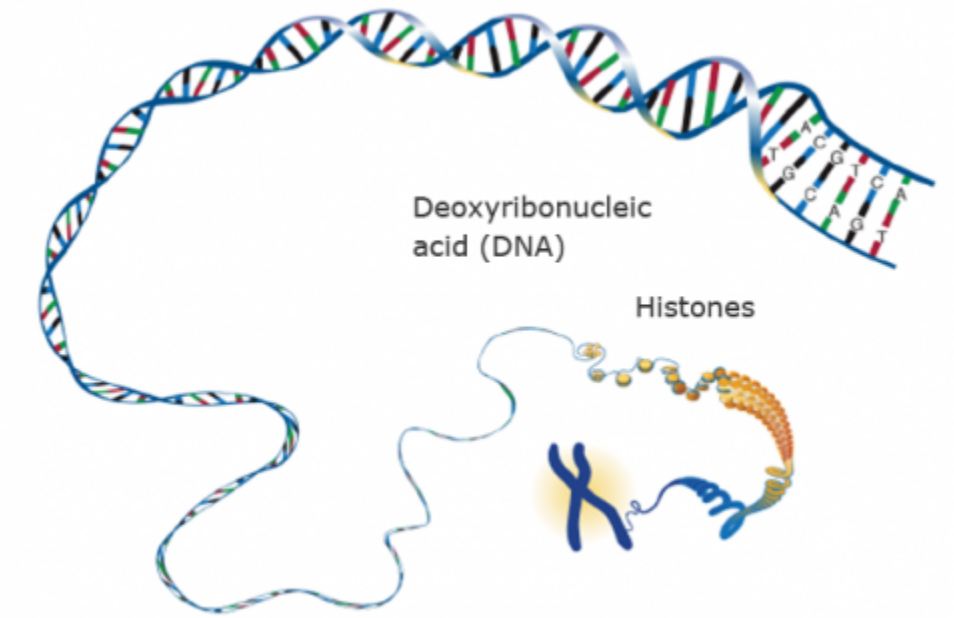


Fig. 2 Nuclear DNA wound around histone proteins that provide structural support to the chromosome. Illustration by NIH-NHGRI.

However, some additional DNA and genetic information is located in specialized cell structures that are "extra-nuclear," meaning found outside of the cell nucleus.

Nuclear vs. Organellar DNA

Although by far the main portion of DNA and genetic information is located on nuclear chromosomes, additional DNA (and genetic information) is located in two types of organelles in plant cells: plastids and mitochondria. Each plant cell usually contains multiple plastids and mitochondria, which contain multiple copies of circular DNA each. Different from the inheritance of nuclear DNA, organellar DNA is maternally inherited in most plant species and not undergoing meiosis.

Mitochondrial DNA is the small circular chromosome found inside mitochondria, a type of organelle found in cells, and that are the sites of energy production.Illustration by NIH-NHGRI.

Chromosomes And Genomes

A **genome** is the basic set of chromosomes inherited as a unit from one parent. **Somatic cells** (non-germ cells) of **diploid** species contain two sets (2n) of the basic genome (**haploid**) (1n) number of chromosomes. Among species, the number of chromosomes varies. Within a species, the chromosome number (2n in somatic cells and 1n in **germ cells**) is ordinarily constant. However, crop species include both diploids and **polyploids**, which are plants with more than two sets of chromosomes in their cells.

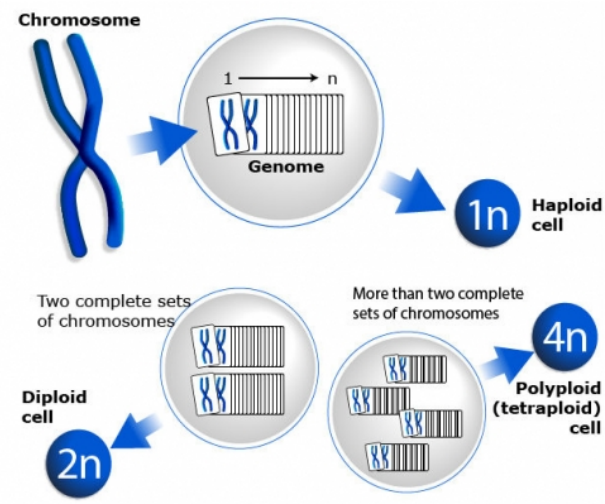


Fig. 3 A complete set of chromosomes forms a genome. Complete genomes are contained in haploid, diploid and polyploid cells.

Crop Chromosome Number

Crop species include a wide range of chromosome numbers. The genomic formula of crop species (with 2n representing the somatic chromosome number, n, the haploid number, and x, the basic chromosome number) can reveal whether or not the crop is a polyploid. For example, vanilla, coconut, pecans, alfalfa, leek, and

sour cherry all have the same number of chromosomes, but the first three are diploids (2n=2x=32), and the last three are polyploids (2n=4x=32). Polyploids can be classified as either autopolyploids (see crop examples labeled as "auto" in the table where their status is known) or **allopolyploids** ("allo" in the table). **Autopolyploids**are polyploids with multiple chromosome sets derived from a single species, whereas allopolyploids are polyploids with genomes derived from different species.

Fig. 4 Crops and their genomic formulas. Photos by Rillke and Alex Popovkin, licensed under CC-SA 3.0 via Wikimedia Commons

Mitosis and Meiosis

Mitosis is one phase of the cell cycle. Mitosis is divided into four arbitrary stages. Progress between stages, however, is gradual and continuous.

During cell division, DNA is duplicated and distributed to daughter nuclei. The number of chromosomes in the daughter nuclei depends on the process of nuclear division. There are two nuclear division processes by which new cells are formed.

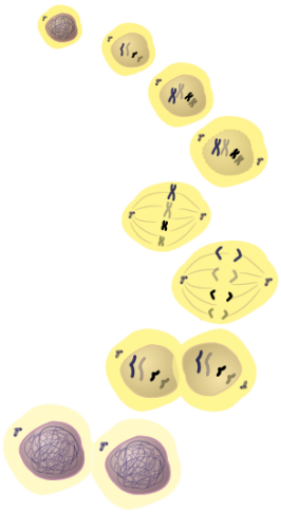


Fig. 5 An encapsulated view of stages in mitosis. Illustration by NIH-NHGRI.

Mitosis

Let's follow mitosis, the process by which somatic cells, non-germ cells, are reproduced.

We'll begin with interphase of the cell cycle. This illustration represents a diploid, or 2n, somatic nucleus. It has two sets of chromosomes, shown here as a blue set and a red set.

Each chromosome is duplicated at the end of interphase during the synthesis period of the cell cycle that precedes mitosis.

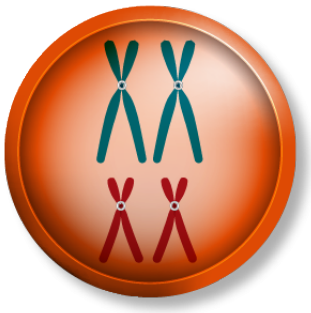


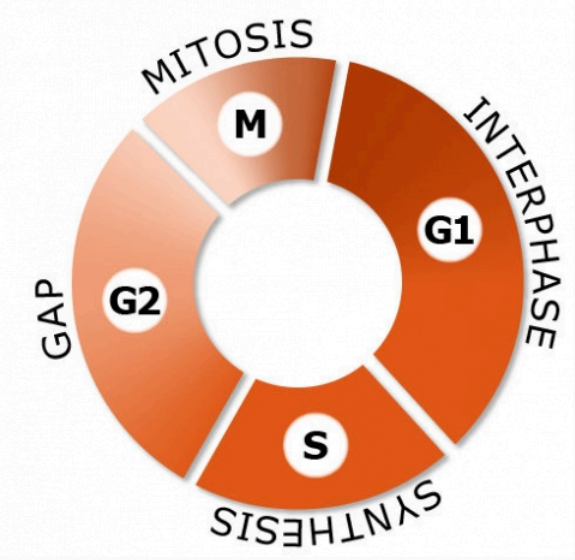
Fig. 6 An illustration of a diploid nucleus with two sets of chromosomes.

Mitosis is a process of nuclear division. It has four stages:

- Prophase
- Metaphase
- Anaphase
- Telophase

Let's see what happens during each of these stages.

FYI: Cell Cycle



- **G2-Gap** This period occurs after DNA replication is complete, but before mitosis begins.
- **M-Mitosis** The nucleus divides, distributing a complete set of chromosomes to each daughter nucleus. The cell subsequently undergoes cytokinesis, cytoplasmic division,

completing the formation of the two daughter cells; each repeats a new cell cycle. Of the cell cycle phases, mitosis is the shortest, typically lasting only 1 to 3 hours.

- **G1-Interphase** Occurs after the completion of mitosis and precedes DNA replication.
- **S-Synthesis** DNA replication occurs. This phase can be identified because it is the only phase during which the cell can incorporate radioactive thymidine into nuclear DNA. Thymidine is related to one of the purine bases of DNA, thymine.

Meiosis

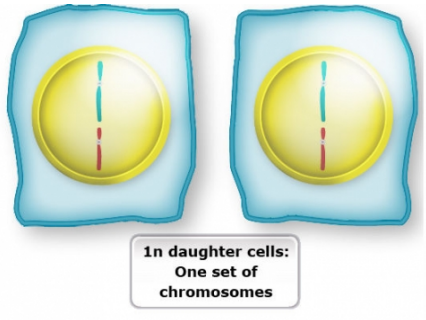


Fig. 7 1n daughter cells: One set of chromosomes.

In contrast to mitosis, meiosis is the process through which germ cells, microspores and megaspores, are derived. Meiosis is similar to mitosis except in two important aspects:

- Meiosis involves two successive divisions.
- Homologous chromosomes replicate only once during the two divisions. Thus, the diploid microspore and mega spore mother cells are meiotically reduced to the haploid, or 1n, chromosome number of the gametes.

In meiosis, there are two successive divisions, called meiosis I and meiosis II. Each of these is divided into four phases analogous to those of mitosis. Like mitosis, these stages progress in a gradual and continuous manner.

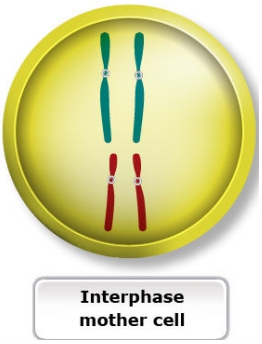


Fig. 8 Interphase mother cell.

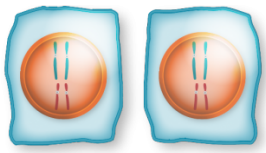
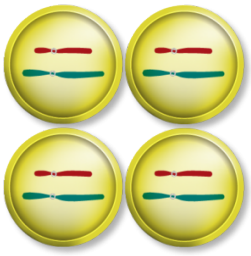
Similarities to Mitosis

Meiosis I resembles mitosis in that:

- The division results in the production of two daughter cells.
- Cells are derived from a microspore or megaspore mother cell.
- Replication of homologous chromosomes precedes it.

Review

Now review these two processes: mitosis and meiosis. Again, pay attention to commonalities and differences. You should be able to identify key features of each process and stage.

Process	 Mitosis	 Meiosis
Divisions	Equational division	One equational division, one reductional division
Results in	Two 2n daughter cells	Four 1n daughter cells
Stages	Four: • Prophase • Metaphase • Anaphase • Telophase	Eight or nine: • Prophase I • Metaphase I • Anaphase I • Telophase I • Interkinesis (sometimes) • Prophase II • Metaphase II • Anaphase II • Telophase II

Sexual Reproduction

Reproduction enables the propagation of new individuals. Reproduction in crop species may occur sexually, asexually, or both.

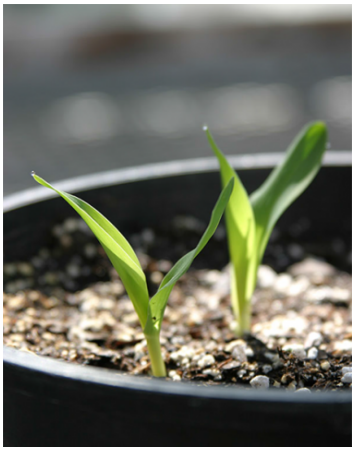


Fig. 9 Reproduction enables new plants to sprout. Photo by Iowa State University.

- **Sexual reproduction**: Requires the fusion of egg and sperm cells (known as **gametes**) to obtain the next generation. The life cycle of a typical angiosperm involves sexual reproduction based on the process of meiosis, in which the chromosome number of cells in the female and male reproductive organs is reduced by half to form female and male gametes. Meiosis is the process responsible for the genetic segregation observed in progeny of **heterozygous** individuals.
- **Asexual reproduction**: Propagation occurs without the fusion of male and female gametes. Asexual reproduction is based on the multiplication of cells by mitosis and results in two new cells that are genetically identical to each other and to the cell from which they originated.

Synopsis of the Life Cycle of the Angiosperm Plant

In the animal kingdom, the production of gametes follows immediately after the meiotic divisions. Normally, therefore, the **gametes** are the only haploid (1n) representatives of the animal life cycle. In plants, however, almost invariably (and without exception in higher plants) the immediate products of the meiotic divisions are not gametes but **spores**. The higher plants (angiosperms) we recognize, e.g., the oak tree, turfgrass, clover, wheat, are the diploid (2n) or **sporophytic** stage of the plant's life cycle. In these plants the haploid vegetative or **gametophytic** stage is short-lived and quite inconspicuous. The sporophyte produces spores as a result of sporogenesis or meiosis. Spores undergo a few nuclear divisions in a process known as gametogenesis to form mature gametophytes. Gametes develop within the gametophytes.

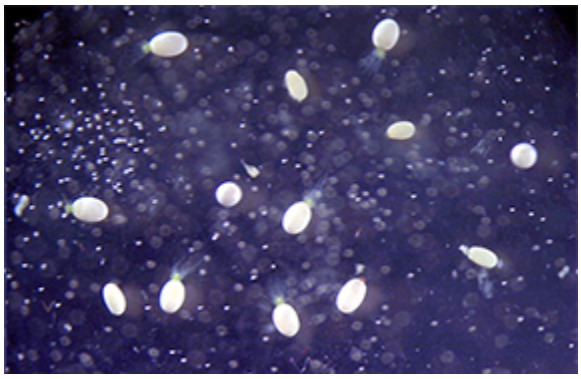


Fig. 10 Microspores and megaspores of Marsilea. Photo by Curtis Clark, licensed under CC-SA 3.0 via Wikimedia Commons.

All higher plants produce two types of spores, **microspores** and **megaspores**. Corresponding to these two types of spores are the two different modes of their development, microgametogenesis and megagametogenesis, which culminate respectively in two dissimilar and relatively simple plants, the mature microgametophytes and megagametophytes.

Male Spore Formation

Each of many 2n microsporocytes (pollen mother cells) within the anther undergoes meiosis with the result that four haploid (1n) microspores are produced within the anther for each original microsporocyte.

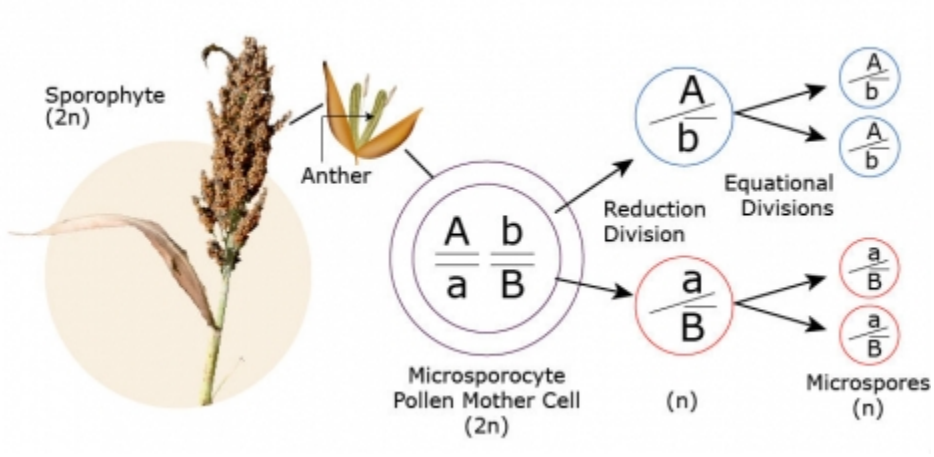


Fig. 11 Microsporogenesis process. Photo by Iowa State University

FYI: Microsporogenesis

Microsporogenesis is the process by which male gametes (pollen grains) are formed. This process can be divided into three parts:

1. Meiosis I
2. Meiosis II
3. Endomitosis

Each of these is subdivided into several stages. The stages of microsporogenesis are transitory. The sequence of stages is shown in the following photos (from Chang and Neuffer, 1989). Each photo represents a momentary expression, which may not be a good representation of the complete event. The event in each photo is indicated by an arrow.

Microgametogenesis

The single nucleus of each microspore divides once by mitosis and one of the two daughter nuclei draws about itself a mass of deeply staining cytoplasm. This nucleus is known as the generative nucleus while the other is the tube nucleus. The generative nucleus undergoes a single mitotic division to form two male gametes, or sperm cells. (In some plants this division does not occur until pollination has taken place and the generative nucleus is moving through the pollen tube.) This constitutes a pollen grain or a mature microgametophyte.

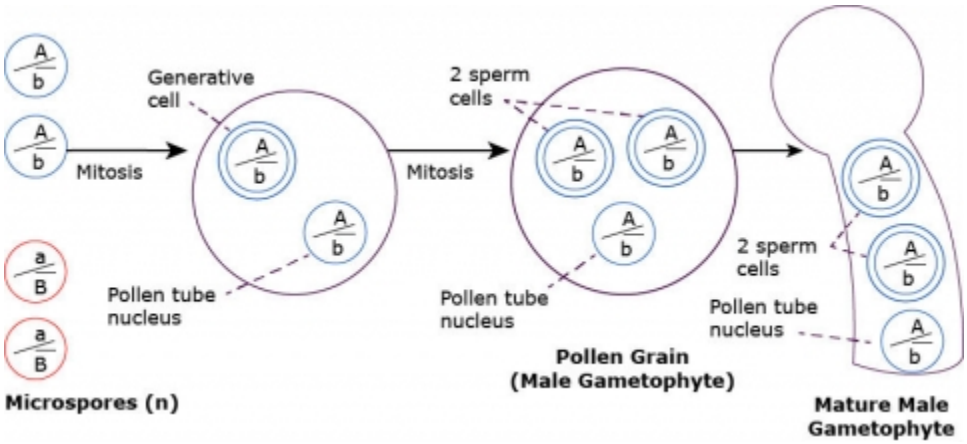


Fig. 12 Microgametogenesis process.

Female Spore and Gamete Formation

Megasporogenesis

A 2n megasporocyte (megaspore mother cell) in each ovule undergoes meiosis. Four megaspores result, each with a haploid chromosome number. Three of these disintegrate; the fourth develops into the mature female gametophyte.

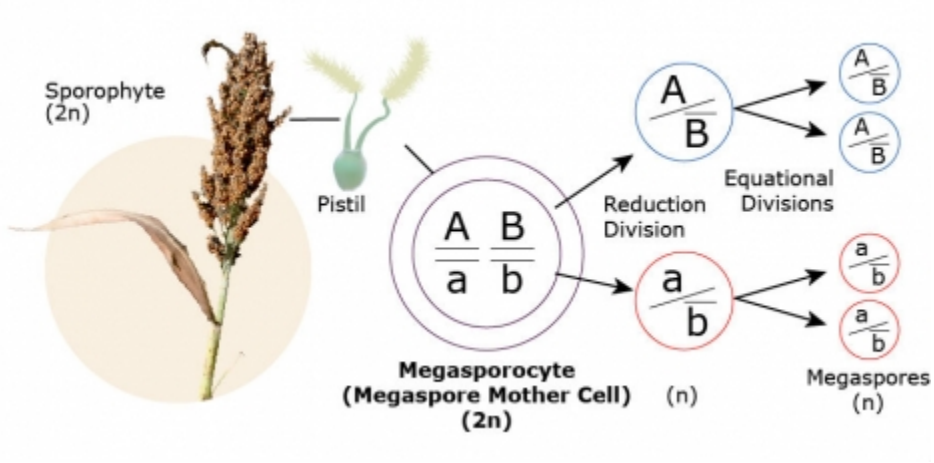


Fig. 13 Megasporogenesis process. Photo by Iowa State University.

Megagametogenesis (development of the female gametophytes and gametes)

The surviving megaspore enlarges greatly to form the embryo sac. Three successive mitotic

divisions, starting with the original nucleus of this megaspore, produce eight haploid daughter nuclei within the embryo sac. These orient themselves as follows:

- The two polar nuclei lie together in the middle of the sac.
- Three nuclei are located at the end of the sac where the sperm will enter, the center one becoming the female gamete or egg, and the two flanking ones the synergids.
- The remaining three, the antipodal cells, come to lie at the opposite end of the sac.

The number of antipodal cells, however varies greatly from zero in Oenothera species to more than 100 in some grass species. The embryo sac with the eight haploid nuclei thus arranged is the mature megagametophyte the female gamete or egg, and the two flanking ones the synergids; the remaining three, the antipodal cells, come to lie at the opposite end of the sac. The number of antipodal cells, however varies greatly from zero in Oenothera species to more than 100 in some grass species. The embryo sac with the eight haploid nuclei thus arranged is the mature megagametophyte.

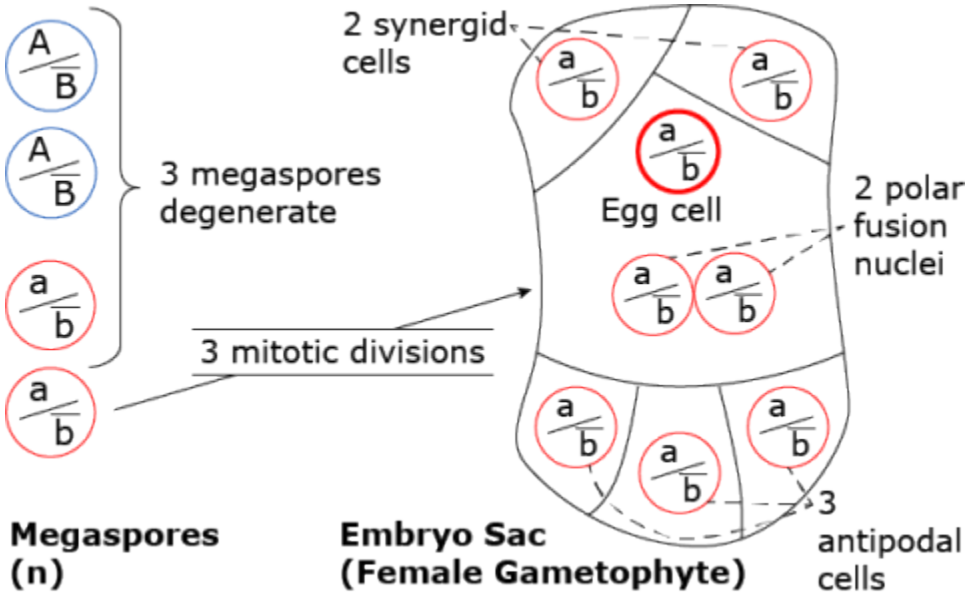


Fig. 14 Megagametogenesis process.

Double Fertilization

Pollen grains are freed by opening of the anther wall and are carried to the stigma of the same

or other plants. Each pollen grain soon sends out a small thin pollen tube (generated by the tube nucleus) which penetrates the tissues of the stigma and digests its way through these and the stylar tissues down to one of the ovules. The sperm cells pass down the pollen tube behind the tube nucleus. Once the pollen tube reaches the ovule it penetrates the embryo sac, the tube nucleus disintegrates, and the two sperm cells enter. One of the sperms fuses with the haploid egg to produce the 2n zygote, while the other fuses with the two polar nuclei to give a 3n (triploid) product, the endosperm nucleus.

Further Development

The triploid endosperm tissue grows more rapidly than the embryo at first. Later on, the embryo, which develops from the zygote, grows at the expense of the endosperm. Depending on the species, endosperm tissue may or may not persist at the time the seed has completed growth. With the resumption of growth (seed germination) the embryo continues development until it reaches the mature sporophyte stage, at which time microspores and megaspores are again produced.

References

Bicknell, A. R. and Koltunow, M., A. 2004. Understanding Apomixis: Recent Advances and Remaining Conundrums [Online]. Available by web link to American Society of Plant Biologists (http://www.plantcell.org/content/16/suppl_1/S228.full.pdf) (verified July 20, 2014).

Bashaw, E.C. 1980. American Society of Agronomy-Crop Science Society of America: 1980 Madison, WI, Hybridization of Crop Plants [Online]. Available at http://www.agron.iastate.edu/faculty/fehr/HOCP/03HOCP.pdf (verified July 20, 2014).

Chang, M.T., and M.G. Neuffer. 1989. Maize microsporogenesis. Genome 32(2): 232-244.

Lersten, N. R. 1980. Reproduction and Seed Development. Chapter 2, p. 17-44. In W.R. Fehr. and H. H. Hadley, editors. Hybridization of Crop Plants. American Society of Agronomy, Madison, WI.

National Institutes of Health. National Human Genome Research Institute. “Talking Glossary of Genetic Terms.” http://www.genome.gov/glossary/

Ozias-Akins, P. 2006. Apomixis: Developmental Characteristics and Genetics. Critical Reviews of Plant Sciences. 25: 199-214.

Pierce, B. A. 2008. Genetics: A Conceptual Approach. 3rd edition. W.H. Freeman, New York.

Schlegel, R.H.J. 2010. Dictionary of Plant Breeding. 2nd edition. CRC Press, New York.

Smartt, J., and N. W. Simmonds. 1995. Evolution of Crop Plants. 2nd. edition. Longman, Essex, UK.

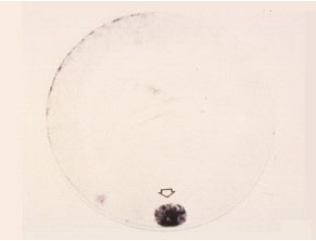
Vielle-Calzada, J-P., C. F. Crane and D. M. Stelly. 1996. Apomixis: the asexual revolution. Science 274: 1322-1323.

mcar2010. 2010. Plant reproduction.mov [Online]. YouTube. Available at http://www.youtube.com/watch?v=ZvBMQIrlhYw (verified July 20, 2014).

For Your Information

Endomitosis

Interphase occurs between meiosis II and endomitosis. Chromosomes replicate during interphase. Endomitosis is a process of cell division resulting in the production of the pollen grains. Endomitosis is also divided into phases. Figures 23 through 36 show these phases, as well as the gradually increasing accumulation of starch granules in the cell. The starch grains progressively obscure the visibility or resolvability of the cellular structures of the male gametophyte.



Interphase

The cell nucleus is round, condensed and nondifferentiated . Chromosomes replicate during this period.

Fig. 15 Interphase.

First Prophase

The chromosomes condense into short thick threads surrounding the nucleolus.

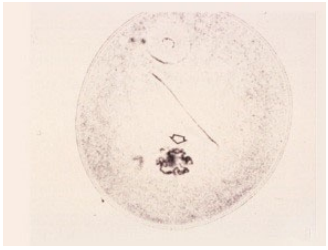


Fig. 16 First Prophase.

Fig. 17 Middle First Prophase.

Middle First Prophase

The chromosomes continue to condense into short thick threads which allow identification of individual chromosomes. The nucleolus and the nucleolus-organizing region of chromosome six are visible.

Late First Prophase

The chromosomes further condense to become short thick rods. The nucleolus and the nucleolus-organizing regions with chromosome six are clearly seen.

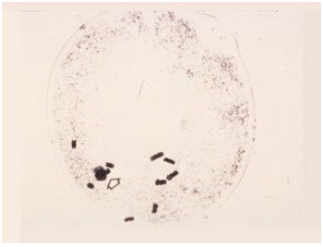


Fig. 18 Late First-Prophase.

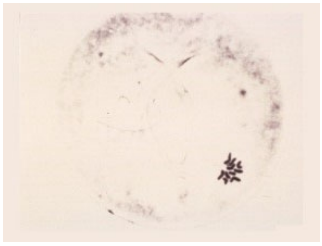


Fig. 19 First Metaphase.

First Metaphase

The nucleolus disappears and the ten chromosomes are arranged in one plane close to one another.

First Anaphase

The sister chromatids are now separated and moving towards the opposite poles.

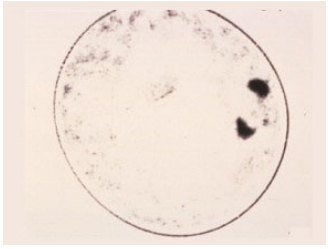
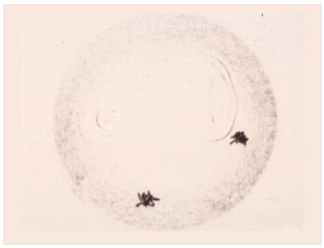


Fig. 20 First Anaphase.



Late First Anaphase

The separated chromosomes have reached the opposite poles and formed two chromosome clusters.

Fig. 21 Late First Anaphase.

First Telophase

The chromosomes at each pole are now extended and surround the nucleolus.

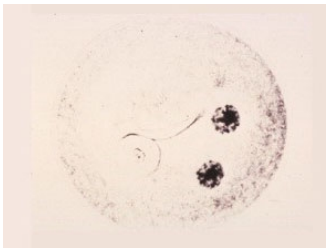
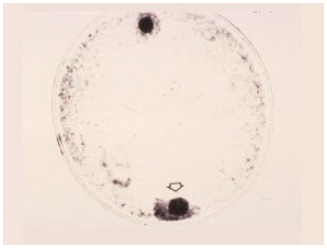


Fig. 22 First Telophase.



Binucleate

Two nuclei are formed at the opposite poles. The generative nucleus (bottom) is usually located near the germ pole. It will further divide to form two sperm.

Fig. 23 Binucleate.

Second Prophase

The generative nucleus underneath the surface of the intine wall becomes cup-shaped and will proceed to the second nuclear division. The vegetative nucleus is not at resting stage and appears to continue its metabolic activities.

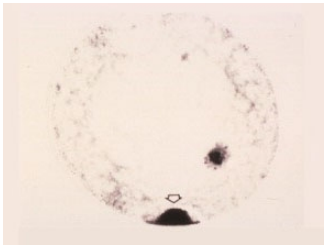
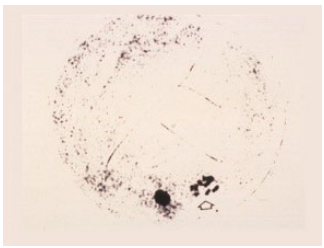


Fig. 24 Second Prophase.



Second Metaphase

The nucleolus disappears, the ten chromosomes are arranged in one single plane. The vegetative nucleus is dark-stained.

Fig. 25 Second Metaphase.

Second Anaphase

The sister chromatids are now separated and moving towards the opposite poles . The vegetative nucleus remains large and clear.

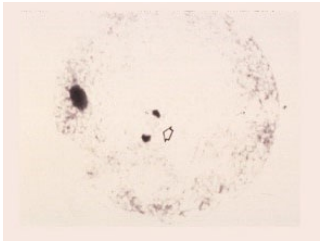


Fig. 26 Second Anaphase.

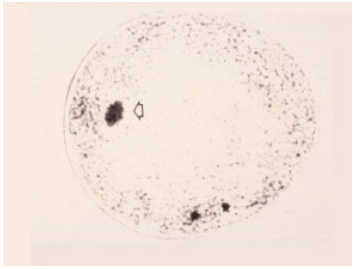


Fig. 27 Second Telophase.

Second Telophase

The chromosomes at each pole are now extended and surround the nucleolus. The vegetative nucleus remains large and clear.

Mature Pollen

The mature pollen grain now has three nuclei. The top two condensed, crescent-shaped nuclei, surrounding the germ pore are the sperms. The large one at the bottom is the vegetative nucleus.

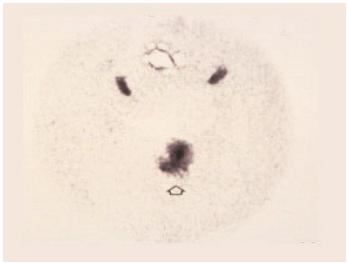


Fig. 28 Mature Pollen.

Meiosis I

Meiosis I is a reductional division-the number of chromosomes in the nucleus is reduced to the haploid number. Meiosis I has four phases: prophase I, metaphase I, anaphase I, and telophase I.

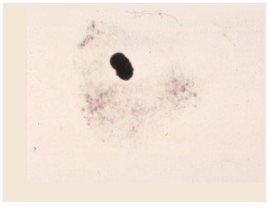


Fig. 31 Premeiotic Interphase.

Premeiotic Interphase

The irregularly shaped pollen mother cell has dense protoplasm, no vacuoles, no clear cell wall structure and an undifferentiated nucleus.

Prophase I

Prophase I is the longest phase of Meiosis I. During Prophase I, the nuclear membrane breaks down, the chromosomes contract, and the spindle forms. Prophase I has several substages: lepotene, zyotene, pachytene, diplotene, and diakinesis.

Leptotene

Cell becomes round with dense protoplasm. The chromatin threads are greatly extended and coiled around the nucleolus. Synapsis is initiated. Single and double strand configuration is evident. The chromomeres are visible.

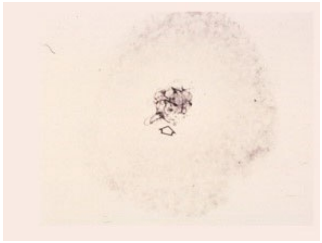


Fig. 32 Leptotene.

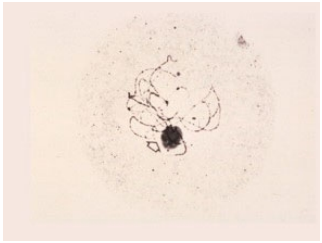


Fig. 33 Late Zygotene-Early Pachytene.

Late Zygotene-Early Pachytene

The pairing of the homologous chromosomes is complete. The condensed chromosomes show details of hetero-chromatin and knobs. The nucleolus and nucleolar-organizing region of chromosome six are visible.

Pachytene

The paired chromosomes are further condensed to become a very thick thread. Individual chromosomes can be identified by their relative lengths, distinctive chromomere patterns, position of knobs, and other recognizable characteristics. The nucleolar-organizing region of chromosome six is clearly attached to the nucleolus.

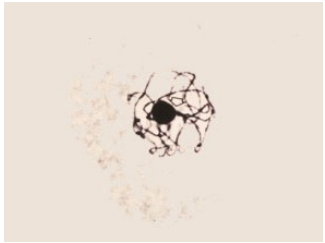
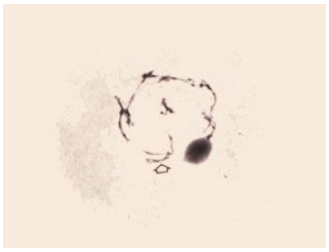


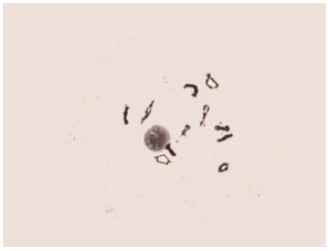
Fig. 34 Pachytene.



The chromosomes continue to condense into short, thick threads. The paired chromosomes appear to be repulsing one another, except regions where an actual crossover took place. The chiasmata are frequently seen as X-shaped and looped chromosome configurations.

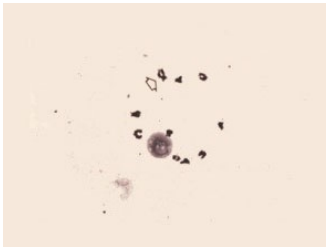
Fig. 35

Late Diplotene



The chiasmata are terminalized and the very short condensed chromosome pairs are separated from each other. The X-shaped and looped chromosome configurations are still shown. The nucleolar-organizing region of chromosome six is firmly attached to the nucleolus.

Fig. 36 Late Diplotene.

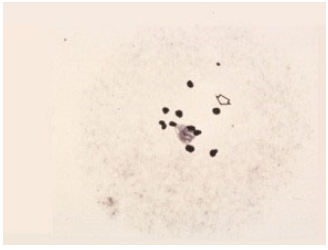


Diakinesis

The condensed chromosome pairs are separated from each other. The chiasmata , the X-shaped and looped configurations are still seen.

Fig. 37 Diakinesis.

Late Diakinesis



The chromosome pairs are dark, round bodies and the nucleolus starts to disappear.

Fig. 38 Late Diakinesis.

Metaphase I

During metaphase I, chromosomes migrate to the spindle equator.

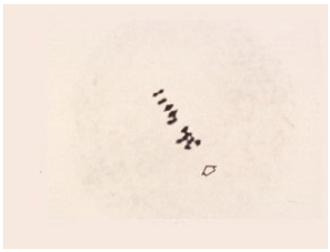


Fig. 39 Metaphase I (side view).

Metaphase I (side view)

The nucleolus has disappeared. The paired chromosomes lie at the equatorial plate of the spindle structure. The chiasmata have moved to the ends of the paired chromosome.

Metaphase I (polar view)

The paired chromosomes appear as dense bodies scattered on a single plane of the protoplast.

Fig. 40 Metaphase I (polar view).

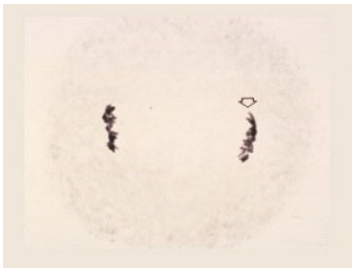


Fig. 41 Anaphase I.

Anaphase I

The paired chromosomes separate and move toward the opposite poles. The V-shaped configuration of the chromosome is due to movement of the centromere ahead of the arms. The number of chromosomes at each pole is now reduced to half the number possessed by the microspore mother cell.

Telophase I

The chromosomes at each pole are now extended. The nucleolus reappears and the cytoplasm divides (cytokinesis) to form two half-mooned cells.

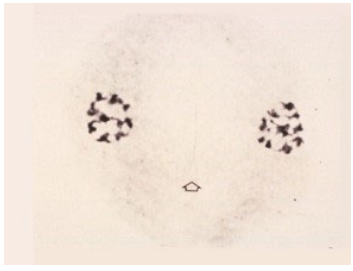
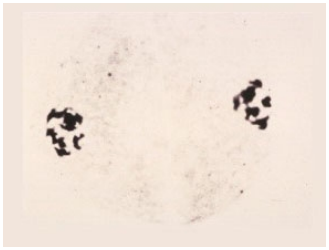


Fig. 42 Telophase I.

Meiosis II

Meiosis II is an equational division during which sister chromatids separate and are distributed to daughter nuclei. Thus, each nucleus receives the haploid number of

chromosomes. Meiosis II is divided into four phases: prophase II, metaphase II, anaphase II, and telophase II. These phases are analogous to the four phases of Meiosis I.



Prophase II

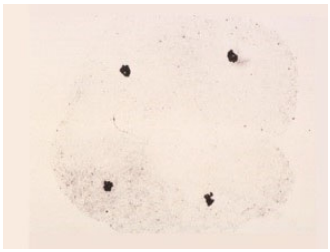
The chromosomes condense into short thick threads surrounding the small nucleolus.

Fig. 43 Prophase II

Metaphase II

The chromosomes (each chromosome has two sister chromatids) lie at the equatorial plate of the spindle structure. Nucleoli have again disappeared.

Fig. 44 Metaphase II.



Anaphase II

The two sister chromatids seen collectively as a dark staining mass, are now separated and have moved towards the opposite poles.

Fig. 45 Anaphase II.

Telophase II

The chromosomes at each pole are extended, the nucleoli reappear and the cytoplasm divides to form four cone-shaped cells.

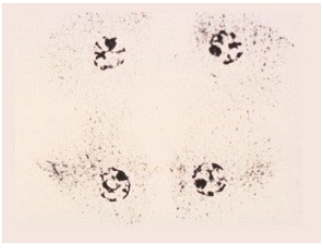
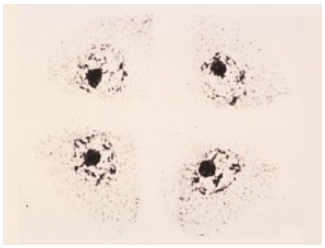


Fig. 46 Telophase II.



Four cone-shaped microspores are formed and are enclosed inside the maternal wall, which is being digested and will thus release the four microspores.

Fig. 47 Telophase II, continued.

The newly released free microspores are undifferentiated, cone-shaped, and appear to have no distinct cell walls.

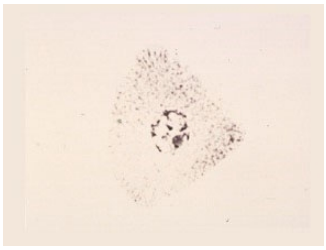
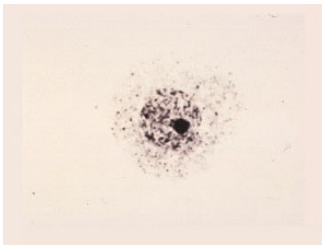


Fig. 48 Free Cell from Quartet.



Early Uninucleate Cell

The shape of the microspores are round with dense cytoplasm. The nucleus is located near the center and the cells are undifferentiated with no vacuoles and no clear wall structure.

Fig. 49 Early Uninucleate Cell.

Late Early-Uninucleate Cell

The microspores start to differentiate. The exine and intine structures are being formed. The cytoplasm remains dense, but many small vacuoles are being formed. The nucleus is still near the center of the protoplast.

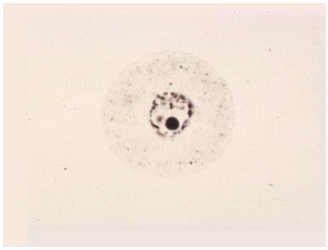
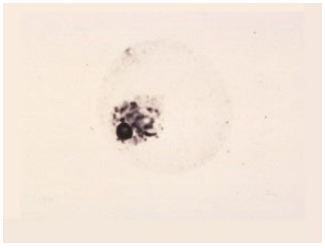


Fig. 50 Late Early-Uninucleate Cell.



Middle Uninucleate Cell

A large vacuole is forming in the protoplast, pushing the nucleus to one side.

Fig. 51 Middle Uninucleate Cell.

Late Uninucleate Cell

The differentiation of exine and intine, germ pore and annulus are complete. Creases seen are due to pressure of coverslip on rigid spherical pollen wall. Cell volume increases four to six times.

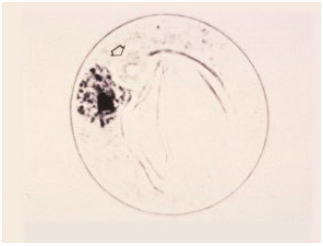


Fig. 52 Late Uninucleate Cell.

How to cite this chapter: Muenchrath, D., A. Campbell, L. Merrick, T. Lübberstedt, & S. Fei. (2023). Reproduction in crop plants. In W. P. Suza, & K. R. Lamkey (Eds.), *Crop Genetics*. Iowa State University Digital Press. DOI: 10.31274/isudp.2023.130

Flower Morphology and Distribution

Deborah Muenchrath; Arden Campbell; Laura Merrick; Thomas Lübberstedt; and Shui-Zhang Fei

Sexual Reproduction

Kinds of Flowers

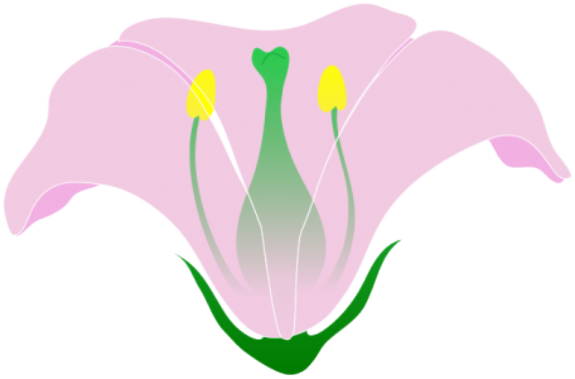


Fig. 1 A complete flower has all four floral organs: Pistil, stamen, sepals and petals.

Fig. 3 This flower is incomplete because it lacks stamens. It is also considered imperfect because it has only one sexual organ (pistil).

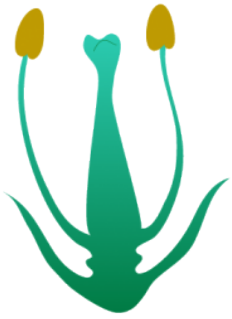


Fig. 2 Incomplete flowers lack one of the four floral organs (here, petals).

Inflorescence type influences the techniques that are used to control pollination in developing cultivars and in maintaining the genetic purity of cultivars. Inflorescence types can also be used to identify plants.

Flowers are classified into a couple of categories. Flowers are either complete or incomplete and either perfect or imperfect. A flower having all of the main floral parts (sepals, petals, pistils, and stamens) is said to be complete, whereas a flower lacking one or more of these structures is said to be incomplete. The stamen (male part) and pistil (female part) are not always present together in a single flower. When both are present, the flower is said to be perfect (or bisexual). Imperfect flowers are those that are unisexual, either male or female.

Table 1 Examples of plants with complete and incomplete flowers.

Complete flowers	**Incomplete flowers**
Soybean	Maize
Alfalfa	Sorghum
Clovers	Oat
Common bean	Barley
Vetches	Wheat
Cotton	Sugar beet
Tomato	**Fig**
Rapeseed	Date palm
Sunflower	Forage grasses
Tomato	Turf grasses
Cabbage	Rice
Tobacco	Spinach

A tomato flower. Licensed under CC-BY-SA 3.0 via Wikimedia Commons.

A fig flower. Licensed under CC-BY-SA 3.0 via Wikimedia Commons.

Notice that plants in the legume family (*Leguminosae* or *Fabaceae*) have complete flowers, whereas plants belonging to the grass family (*Gramineae* or *Poaceae*) have incomplete flowers.

Flower Dissection

Dissect a complete and incomplete flower. Think about how the presence or absence of a floral structure might influence the pollination process, and thus, the methods that can be used to develop improved cultivars or to maintain the genetic purity of the cultivar.

Complete Flower

Fig. 4 Soybean flowers in the field.

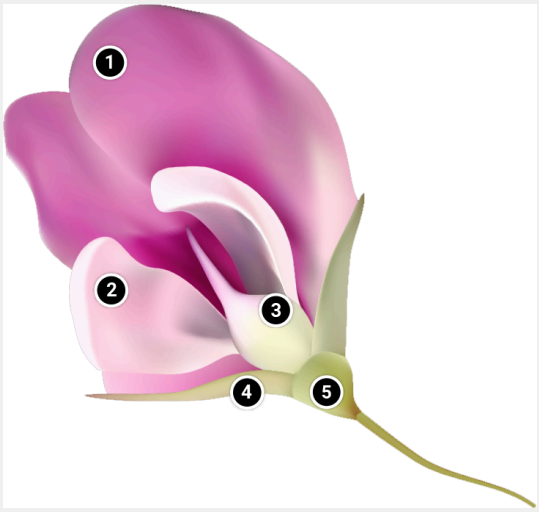
Incomplete Flower

Fig. 5 Green wheat florets. Photo by Stephen C Dickson, CC BY-SA 4.0.

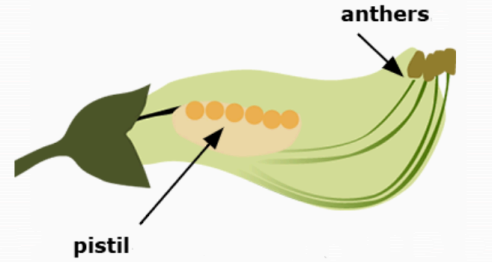
Grass Floret

Fig. 6 Grass flowers. Photo by Hardyplants via Wikimedia Commons

Complete Soybean Flower Dissection



1. **Standard petals:** Collectively, petals are called the corolla. Petals are typically large and conpicuous and are not required for reproduction. Soybean has five petals: one standard petal, two wing petals and two keel petals
2. **Wing petals:** The dissected view of the two wing petals.
3. **Keel:** The keel is composed of two united petals. The keel encloses the stamina column. Stamens are the pollen-bearing organs of the flower. Stamens are composed of slender stalks (filaments) that support anthers.



Pollen grains are produced in the **anthers**. The **pistil** is the seed-bearing organ of the flower. It consists of stigma, style, and ovaries. The stigma is the part that is receptive to pollen. Following pollination and fertilization, seed form in ovaries.

4. **Sepals:** Like the petals, sepals are not neccessary for reproduction. Sepals are small and inconspicuous. They enclose and protect the flower while still a bud. Collectively they form the calyx.
5. **Pedicel:** The pedicel is the stalk of the flower, attaching to the plant.

Table 2 Examples of crops and different floral systems. Adapted from Lersten (1980).

Flower Characteristics	Terms	Examples
Male and female expression in INDIVIDUAL FLOWERS		
Male and female in ONE flower	bisexual, hermaphroditic, monoclinous, perfect	Wheat, peach
1. Pollen shed before stigma is receptive	protandry (prevent self-pollination)	Carrot, walnut
2. Stigma matures and ceases to be receptive before pollen is shed	protogyny (prevent self-pollination)	Pearl millet, pecan
3. Stigma receptive, and pollen shed, after flower opens	chasmogamy (promote self-pollination)	Violet, rye
4. Stigma receptive, and pollen shed, in closed flower	cleistogamy (ensure self-pollination)	Oat, peanut
Perfect flowers of TWO types on SAME plant	heterostyly	Buckwheat, flax
1. Long styles and short stamens	pin flower	
2. Short styles and long stamens	thrum flower	
Male and female in SEPARATE flowers	unisexual, diclinous, imperfect	Cucumber, hemp
1. Male flower	male, staminate	
2. Female flower	female, pistillate, carpellate	
Flower DISTRIBUTION on PLANTS		
Male and female flowers on one plant	monoecious	Maize, oak

Flower Characteristics	Terms	Examples
Male and female flowers on separate plants	dioecious	Yams, asparagus
• Male, female, and perfect flowers	mixed, polygamous	Red maple, papaya
1. On same plants	polygamomonoecious	Coconut, mango
2. On separate plants	polygamodioecious	Strawberry, holly

Perfect and Imperfect Flowers

Perfect flowers have both staminate and pistillate structures in the same flower.

Imperfect flowers are either staminate or pistillate. An imperfect flower is staminate if it possesses stamen. Conversely, an imperfect flower is pistillate if it bears a pistil. Staminate flowers are considered "male" because they produce pollen, whereas pistillate flowers are "female" because they possess ovules. Staminate and pistillate flowers may occur on the same or different plants of the same species.

Species having such specializations are either:

- **monoecious** — staminate and pistillate flowers are separate but occur on the same plant; or
- **dioecious** — staminate and pistillate flowers are on separate plants.

Analogous to the separate sexes in animals, a dioecious plant must have a partner of the opposite type to complete its life cycle. Usually, about half of all individuals of a dioecious species are of each type, staminate or pistillate. Thus, the dioecious condition is reproductively expensive in that only about half of the species' plants can produce seed.

Table 3 Examples of monoecious and dioecious plants.

Monoecious		Dioecious	
Maize		Hemp	
Walnut		**Hops**	
Oil palm		Spinach	
Squash		Yam	
Cassava		Date palm	
Wile rice	Walnut flower. Licensed under CC-BY-SA 3.0 via Wikimedia Commons.	Cottonwood	Hops flowers. Licensed under CC-BY-SA 3.0 via Wikimedia Commons.
Castor bean		Asparagus	
White pine		Nutmeg	

The "mono-" prefix indicates one and the "di-" prefix indicates two. The "-oecious" part of the word translates to "house." Thus, an easy way to remember the distinction between these terms is to remember that in monoecious species, the staminate and pistillate flowers reside in the same house or plant, whereas in dioecious species, these flowers reside in two different houses or plants.

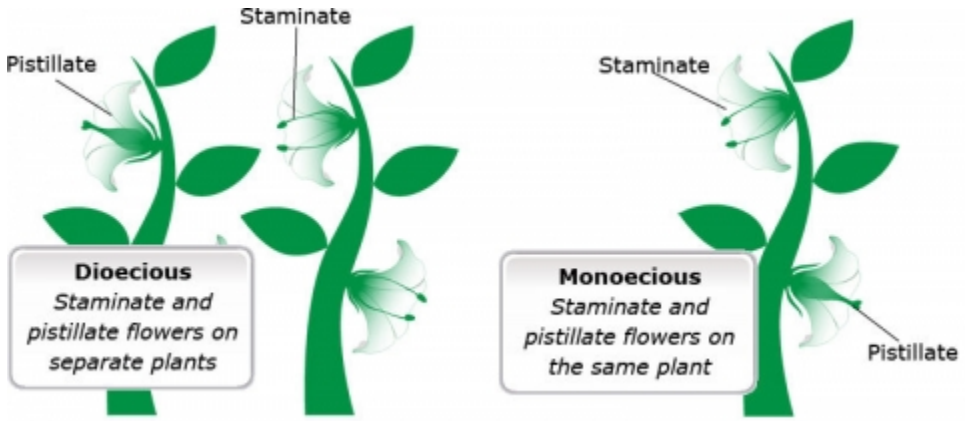


Fig. 7 Dioecious and monoecious plants.

Pollination and Fertilization

Pollination occurs when a pollen grain (from the staminate flower) is placed on a receptive stigma (of the pistillate flower), either naturally or artificially. Fertilization requires that a male

gamete and a female gamete fuse to form a zygote. These gametes may be from the same or different plants.

There are two kinds of pollination processes in sexual reproduction.

- **Self-pollination** — seeds develop from the union of male and female gametes produced on the **same** plant or clone. The development of seed by self-pollination is also referred to as **autogamy**.
- **Cross-pollination** — seeds develop from the fusion of gametes produced on **different** plants. The development of seed by cross-pollination is known as **allogamy**.

Self-Pollination

Several floral mechanisms enforce self-pollination.

- Flowers do not open, preventing external pollen from reaching the stigma.
- **Anthesis** occurs before the flower opens.
- Stigma elongates through the staminal column (filaments and anthers) immediately after anthesis.
- Floral organs may obscure the stigma after the flower opens.

Although these mechanisms usually enforce self-pollination, a low frequency of cross-pollination may occur. The frequency of cross-pollination in normally self-pollinating species generally depends on the species and environmental conditions.

Soybean is an example of a species that is normally self-pollinated. Before the flower opens, the anthers burst and pollen grains fall out of the anthers on to the receptive stigma contained in the same flower: self-pollination occurs.

Cross-Pollination

Floral Mechanisms of Promotion

Several floral mechanisms promote cross-pollination.

- Emergence or maturity of the staminate and pistillate flowers is asynchronous.
  - **Protandry** — anthesis occurs before stigma are receptive.

    - **Protogyny** — pistillate flower matures before the staminate flower.

- Flowers are monoecious or dioecious. Mechanical obstruction between the staminate and pistillate flowers in the same individual prevents self-pollination. Gametes produced on the same plant or clone are unable to effect fertilization.

- Mechanical obstruction between the staminate and pistillate flowers in the same individual prevents self-pollination.

Alfalfa flowers, for example, have a membrane over the stigma that precludes self-pollination. When a bee lands on the flower, the keel is tripped, rupturing the membrane and exposing the stigma to pollen carried by the bee from other plants it has visited, effecting cross-pollination.

Fig. 8 An alfalfa leafcutting bee on an alfalfa flower.

- Gametes produced on the same plant or clone are unable to effect fertilization.
    - **Self-sterility** — gametes from same individual cannot successfully fuse to form a zygote. Sterility can be caused by lack of function of pollen (male gametes) or ovules (female gametes).
    - **Male sterility** — either genetic or cytoplasmic, occurs because the pollen is not viable. Female sterility occurs when the ovule is defective or seed development is inhibited.
    - **Self-incompatibility** — self-pollination may occur, but fertilization and seed set fail.

Pollen Transportation

Pollen is transported from the staminate flower to the pistillate flower by wind, insects, or animals. Occasionally pollen is transported to receptive stigma of the same individual and self-pollination may occur. For example, pollen from the tassel of a maize plant may land on and pollinate silks on the same plant, effecting self-pollination.

Sunflower is ordinarily cross-pollinated. Bees often carry pollen from one plant and deposit it on other plants.

Fig. 9 Close-up of a sunflower head. Photo by Allen Knapp, Iowa State University.

Fig. 10 Beehives on the edge of a sunflower field. Photo by Allen Knapp, Iowa State University.

Classification

Plants are classified as either self- or cross-pollinated based on which of these processes most frequently produces its seed. Click each category for more information.

Asexual Reproduction

Some species can be propagated without a gametophytic stage. The fusion of gametes (fertilization) is omitted from the life cycle. Reduction in chromosome number (meiosis) and seed production may or may not occur. Asexual reproduction produces individuals genetically identical to the maternal parent.

There are several mechanisms of asexual reproduction.

- Vegetative Propagation

- Tissue Culture
- Apomixis

Vegetative Propagation

In some species, new individuals can arise from a group of differentiated or undifferentiated cells of the parent plant; no embryo or seed is produced. Because such new individuals develop asexually from a single parent, they are genetically identical to that parent. These progeny are **clones**. Numerous tissues and organs may asexually produce progeny.

- **Rhizomes** – Rhizomes are specialized underground stems that can branch at nodes to produce new plants. Banana, bromegrass, hops, and johnsongrass can be reproduced from rhizomes.
- **Stolons** – These "runners" or horizontal-growing, above-ground stems develop adventitious roots whose axillary buds can become independent plants. Strawberry is an example of a crop that can be reproduced from stolons.
- **Bulbs and bulbils** – These short underground stems have thickened or fleshy scales (modified leaves) that can form buds. These buds detach and form "offsets" or new individuals. Onions and garlic are commonly propagated from bulbs.
- **Tubers** – Tubers are also short, enlarged stem tissue, containing food reserves. Nodes or "eyes" in such tissue can give rise to adventitious roots and separate plants. Potatoes are commonly propagated from eyes cut from tubers.
- **Suckers** – Suckers arising as lateral shoots from the base of stems can separate and form new plants. Pineapple, sweet potato, and date palm are examples. Suckers may also derive from adventitious buds on the roots. Roses, poplars, and some other woody species can be propagated from such root cuttings or rootstocks.
- **Corms** – A corm is an underground, tuber-like base of a vertical stem that can also produce a separate plant. Taro, an important starch crop in Southeast Asia and the Pacific Islands, is propagated from corms. Banana also can be propagated from corms.
- **Stem cuttings** – When placed in moist soil, cuttings from aerial stems of some species, such as sugarcane, pineapple, and cassava, can give rise to new plants from the nodes and lateral buds.

The usual mode of reproduction of some species is vegetative. However, other species that

reproduce sexually are more commonly propagated vegetatively to maintain genetic purity, including some forage cultivars and many horticultural species.

Vegetative reproduction does not usually provide opportunity for selection of genetic variants.

Tissue Culture

Tissue culture is a specialized type of asexual propagation. Tissue culture usually involves excision of undifferentiated cells or **meristematic** pieces of a plant and growing these in vitro on sterile nutrient agar medium; cell division is by mitosis. By manipulating the components of the medium, the tissue can be prompted to develop roots or shoots. Eventually, new individuals may be separated and transplanted to soil.

Tissue culturing takes advantage of the **totipotency** of somatic cells. That is, these cells contain the plant's entire genome and have the potential to develop into whole plants. Some species that cannot normally be reproduced vegetatively may be reproduced by tissue culture.

Tissue culture is of interest to plant breeders as a technique to

- maintain and propagate genetically identical plants that otherwise can only be reproduced sexually;
- provide disease-free plants of species that often transmit pathogens to progeny when propagated by conventional vegetative means; and
- create novel genetic variation within which selections can be made. Under some conditions, tissue culturing can promote genetic changes.

Apomixis Process

Apomixis differs from other forms of asexual reproduction in that seed is produced. Unlike sexual reproduction, however, apomictic seed is developed from sexual organs or related structures without fertilization. Pollination is also usually omitted.

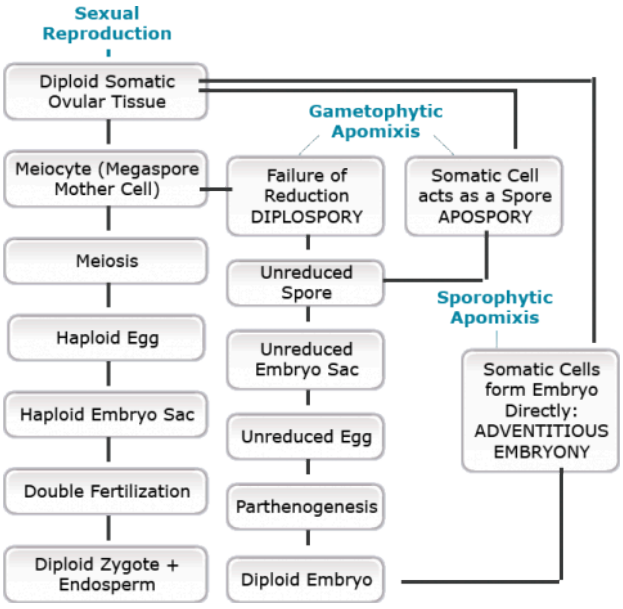


Fig. 11

Agamospermy

Apomixis generally involves forms of **agamospermy**, which is a process through which seeds develop without fertilization. There are two different degrees of agamospermy.

- **Obligate** — Seed produced arises from asexual reproduction.
  - **Advantages:** Preserves genotype, including heterozygotic genotypes
  - **Disadvantages:** Precludes genetic recombination and variation for selection of improved cultivars
- **Facultative** — Although most of the seed generated is asexually produced, sexual reproduction occurs regularly.
  - **Advantages:** Permits development of genetic variation for selection of improved cultivars
  - **Disadvantages:** Cultivars may be genetically unstable, making it difficult to maintain the desired genotype

Each of these degrees of agamospermy provides advantages and disadvantages from the plant breeding perspective.

There are also two general types of agamospermy.

- **Autonomous** — Endosperm forms without pollination or fertilization.
- **Pseudogamous** — Although fertilization (the fusion of gametes) does not occur, pollination is apparently required to stimulate apomictic embryo or embryo sac development to produce seed. Pollination adds no genetic material.

Mechanisms of Cause

The mechanisms that cause apomixis differ by the cell that undergoes mitosis to produce the embryo of the seed.

- **Adventitious embryony** — The embryo develops directly from diploid sporophytic tissue, skipping the gametophytic stage. This is the simplest form of agamospermy.
- **Apospory** — Nucellus or integument cells, which are somatic cells, undergo mitosis to produce a diploid embryo sac.
- **Apospory** is the most common form of apomixis in angiosperms.
- **Diplospory** — The embryo and endosperm derive from the diploid megaspore mother cell. The megaspore mother cell's nucleus divides by mitosis, rather than meiosis, resulting in a diploid embryo sac.
- **Parthenogenesis** — The egg cell divides mitotically to form the embryo without fertilization.
- **Androgenesis** — A haploid embryo develops from a male sperm nucleus after it enters the embryo sac. The individual that develops from the seed is haploid and has the genotype of the sperm from which it is derived.

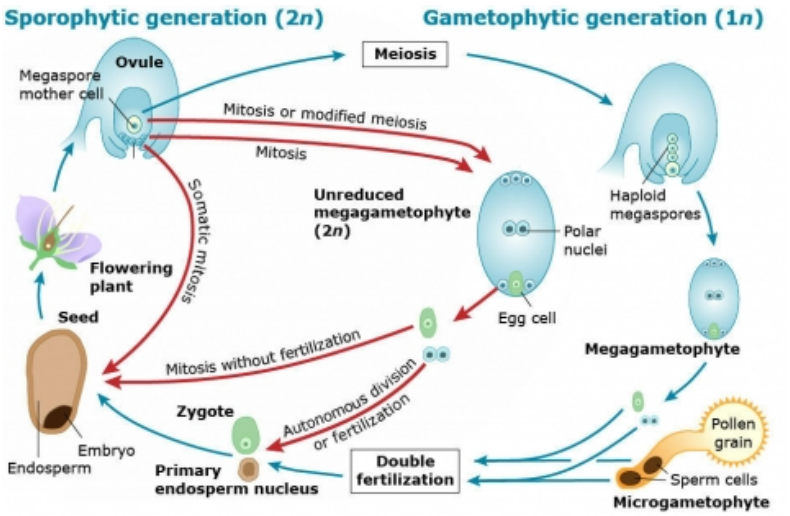


Fig. 12 The sexual life cycle of flowering plants and the apomictic reproduction by seed which occurs when the sexual life cycle is short-circuited. Adapted from Vielle-Calzada et al., 1995.

Discussion

Crops can be self-pollinated, cross-pollinated, or vegetatively propagated. Discuss the breeding consequences of these three different methods of propagation. In addition: previously, a student suggested that with today's technologies, plants can simply be converted into self- or cross-pollinated or into vegetatively propagated species. Do you agree? Provide arguments in favor or against this statement, and examples, in case you are aware of any. Finally, if it was possible, which type of crops would be your favorite, and why?

References

Bicknell, A. R. and Koltunow, M., A. 2004. Understanding Apomixis: Recent Advances and Remaining Conundrums [Online]. Available by web link to American Society of Plant Biologists (http://www.plantcell.org/content/16/suppl_1/S228.full.pdf) (verified Jan 24, 2017).

Bashaw, E.C. 1980. American Society of Agronomy-Crop Science Society of America: 1980 Madison, WI, Hybridization of Crop Plants [Online]. Available at http://www.agron.iastate.edu/faculty/fehr/HOCP/03HOCP.pdf (verified Jan 24, 2017).

Chang, M.T., and M.G. Neuffer. 1989. Maize microsporogenesis. Genome 32(2): 232-244.

Lersten, N. R. 1980. Reproduction and Seed Development. Chapter 2, p. 17-44. In W.R. Fehr. and H. H. Hadley, editors. Hybridization of Crop Plants. American Society of Agronomy, Madison, WI.

Ozias-Akins, P. 2006. Apomixis: Developmental Characteristics and Genetics. Critical Reviews of Plant Sciences. 25: 199-214.

Pierce, B. A. 2008. Genetics: A Conceptual Approach. 3rd edition. W.H. Freeman, New York.

Schlegel, R.H.J. 2010. Dictionary of Plant Breeding. 2nd edition. CRC Press, New York.

Simmonds, N. W., and J. Smartt. 1999. Principles of Crop Improvement. 2nd. edition. Blackwell Science, Oxford, UK.

Smartt, J., and N. W. Simmonds. 1995. Evolution of Crop Plants. 2nd. edition. Longman, Essex, UK.

Vielle-Calzada, J-P., C. F. Crane and D. M. Stelly. 1996. Apomixis: the asexual revolution. Science 274: 1322-1323.

mcar2010. 2010. Plant reproduction.mov [Online]. YouTube. Available at http://www.youtube.com/watch?v=ZvBMQIrlhYw (verified Jan 24, 2017).

How to cite this chapter: Muenchrath, D., A. Campbell, L. Merrick, T. Lübberstedt, & S. Fei. (2023). Flower morphology and distribution. In W. P. Suza, & K. R. Lamkey (Eds.), *Crop Genetics*. Iowa State University Digital Press. DOI: 10.31274/isudp.2023.130

Controlled Hybridization: Self-incompatibility, Male-sterility, and Sex-inheritance

Deborah Muenchrath; Arden Campbell; Laura Merrick; Thomas Lübberstedt; and Shui-Zhang Fei

Introduction

Most crop species require seed production for their propagation. Some species, however, possess mechanisms that regulate fertility. Such mechanisms can reduce or prevent seed set, and affect self- or cross-pollination. These fertility-regulating mechanisms may be an obstacle or a benefit to the plant breeder. In this module, we'll explore these mechanisms and their utility.

Learning Objectives

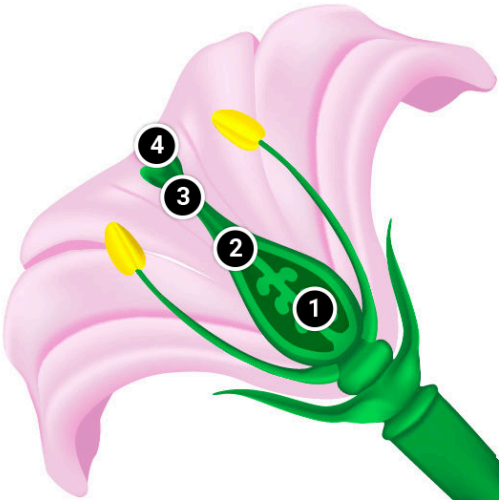
- Mechanisms and utility of incompatibility systems.
- Modes of sex inheritance in plants and their application in plant breeding.
- Male sterility systems and their applications in plant breeding.

Self-incompatibility

Genetic-based Self-incompatibility

Self-incompatibility is the inability of a plant to set seed when self-pollinated, even though it produces viable pollen. In contrast, cross-pollination generally results in seed set in self-incompatible species. Many plant families include species with self-incompatibility systems, such as *Fabaceae* (*Leguminosae*), *Poaceae* (*Gramineae*), *Solanaceae*, *Brassicaceae* (*Cruciferae*), and

Asteraceae (*Compositae*). Self-incompatibility may be caused by genetic interactions between pistil and pollen, or by physical obstacles that hinder self-fertilization.



Self-incompatibility may be caused by genetic interactions between the pistil and pollen-producing physiological factors that interfere with fertilization of female gametes by male gametes produced on the same plant or on a closely-related plant. Typically one or few self-incompatibility (SI) genes are involved in this self/non-self discrimination process, depending on the plant species. Pollen can be rendered ineffectual at several points in the pollination process:

1. Sperm enters the embryo sac but does not fuse with the egg.
2. Pollen tube penetrates the stigma but it grows too slowly in the style to reach the ovary while the ovule is still receptive.
3. Pollen germinates but the tube is unable to penetrate the stigma.
4. Pollen germination on the stigma is inhibited.

Pistil receptivity to a particular pollen grain depends on the **SI** alleles carried by the pistil. The phenotype of the pollen (its capacity to fertilize the female gamete) is determined by either the pollen's alleles (gametophytic incompatibility) or by the alleles of the plant that produced the pollen (sporophytic incompatibility).

There are two self-incompatibility systems that result from genetic interactions:

- Gametophytic Self-incompatibility
- Sporophytic Self-incompatibility

Both incompatibility types influence the rate of pollen tube growth, but their genetic controls and location of effect differ. Both types involve multiple alleles.

Gametophytic Self-incompatibility

- Gametophytic self-incompatibility involves the allele possessed by the pollen grain. The

incompatibility effect occurs in the style. However, in some species with gametophytic SI, incompatibility is expressed on the surface of stigma (in grasses).

- Gametophytic self-incompatibility is controlled by a series of alleles. The rate of pollen tube growth responds to the allelic interaction of both the style and the pollen.
- If both the stylar tissue and pollen possess identical alleles, pollen tube growth is inhibited.
- If stylar and pollen alleles differ, tube growth occurs at normal rates.

FYI: Homozygous Flowers

What happens if the pistillate flower is homozygous for the S allele?

The same rules apply: if the pollen carries an allele that matches the one possessed by the pistil, fertilization will not occur; if they differ, seed can form. However, it should be noted that plants homozygous for a self-incompatibility allele in the gametophytic system are rare.

Why are plants homozygous for gametophytic self-incompatibility rare?

Homozygotes are unusual because the probability of a pollen grain carrying the same allele as the pistil successfully overcoming the incompatibility of their matching alleles is small—but such events do occur occasionally. Thus, a population may have a few individuals homozygous for an S allele.

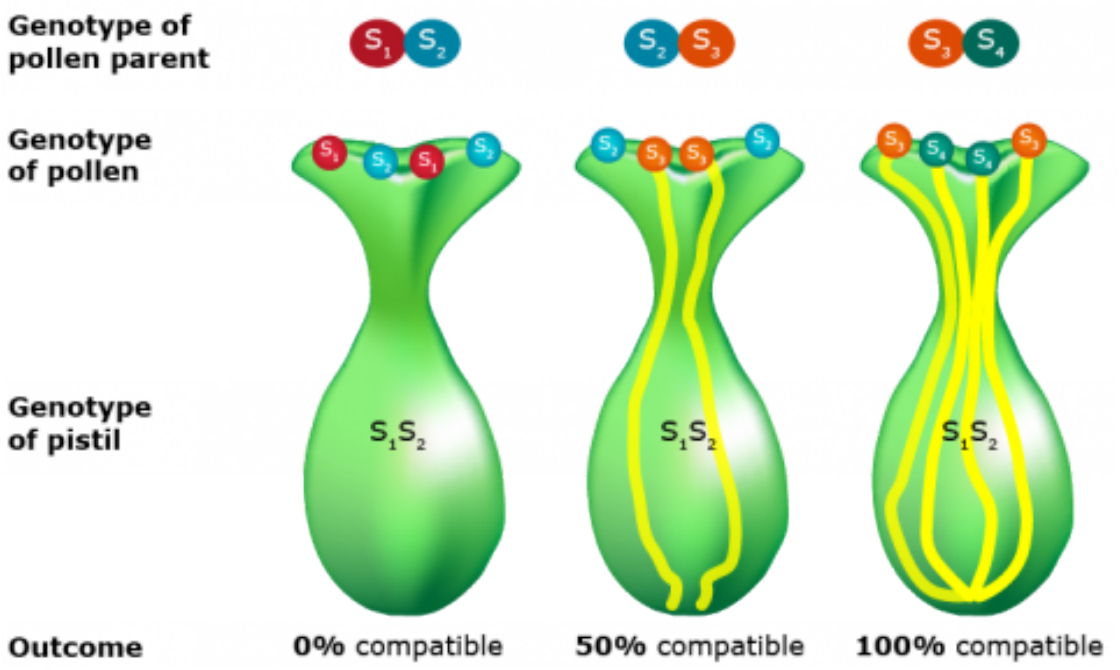


Fig. 1 The working model for gametophytic SI controlled by a single multiple-allelic loci S. When a pollen grain carries an S allele that is matched in the pistil, incompatibility occurs, resulting in pollen tube growth arrest. Otherwise, pollen is allowed to grow into the ovule, where it may participate in fertilization.

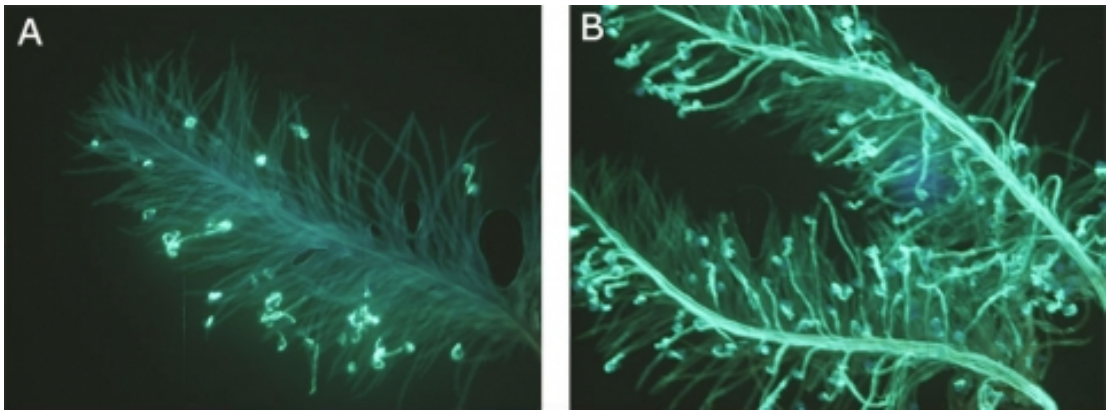


Fig. 2 Pollen-pistil interaction in Miscanthus Sinensis, a warm-season perennial grass that has potential for biofuel production. Photos by Iowa State University.

In Fig 2A, all pollen grains are incompatible with the pistil as they fail to germinate beyond the stigmatic surface.

In Fig 2B, nearly all pollen grains are compatible with the pistil as they germinated and pollen tubes grew through the stylar tissues and eventually entered the ovule for fertilization.

Sporophytic Self-incompatibility

Sporophytic self-incompatibility involves dominance and depends on the allelic composition of the plant that produced the pollen. Incompatibility is expressed at the surface of stigma.

In sporophytic self-incompatibility systems, the rate of pollen tube growth depends on the presence or absence of a dominant allele at the **SI** locus (will be called **S** locus in the following) carried by the pollen-producing plant. In the pollen, any **S** allele can exhibit dominance- that dominance is determined by the sporophyte, the plant that produces the pollen (thus, this system of self-incompatibility is termed "sporophytic"). Two important points need to be emphasized about sporophytic self-incompatibility.

The genotype of the pollen-producing plant determines self-incompatibility, not the allelic composition of the pollen itself. In other words, every male gamete has the same ability to fertilize a female as every other male gamete, irrespective of the pollen's individual genotype.

There is no dominance in the female stigmatic tissue.

How does the genotype of the pollen parent transmit the influence of the dominant allele to the pollen?

Although we don't yet understand this, we describe it as 'imprinting.' That is, the pollen "remembers" the genetic environment in which it developed and is conditioned to behave in accordance with that environment. Cell walls of pollen grains are consisted of at least two layers, the intine or inner layer and the exine or the outer layer. The exine is made up of a highly durable organic polymer, sporopollenin. The exine is believed to be derived from the somatic tissues of the pollen-producing parent and likely play a role in pollen-pistil interaction. This could explain why the compatibility response is determined by pollen-producing parent, instead of the pollen itself.

Here's an example of sporophytic self-incompatibility. Assume the species is diploid and that there are four possible self-incompatibility alleles: $\mathbf{S_1}$, $\mathbf{S_2}$, $\mathbf{S_3}$, $\mathbf{S_4}$. Let $\mathbf{S_1}$ be a dominant allele. Let's see what happens. (Color indicates the source of the allele, from the female or the male parent.)

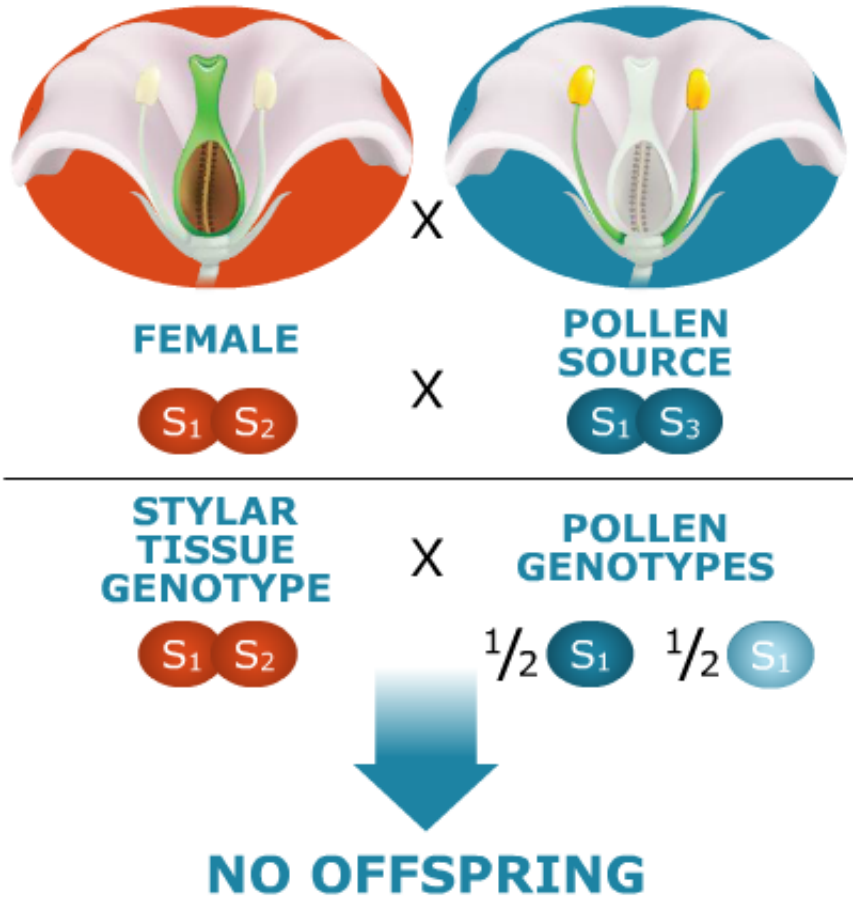


Fig. 1 Animation demonstrating sporophytic self-incompatibility

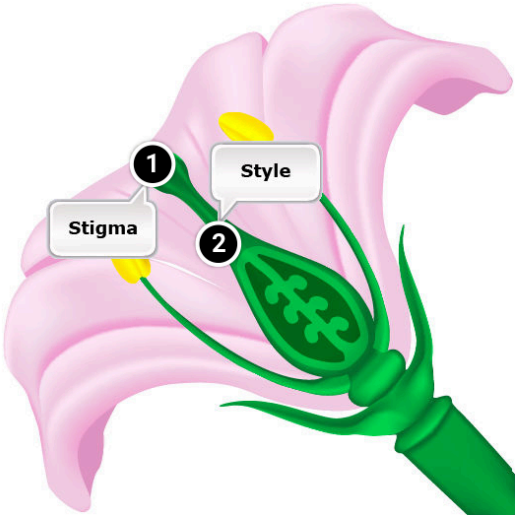
Why are no offspring produced?

The $\mathbf{S_1}$ allele in the plant producing the pollen is dominant. Hence, both the gamete types produced by that plant, $\mathbf{S_1}$ and $\mathbf{S_3}$, will behave as if they were both dominant alleles ($\mathbf{S_1}$). As a result, neither pollen type will be able to effect fertilization of this female—neither type of pollen tube will be able to penetrate the style because their growth will be impeded on the $\mathbf{S_1S_2}$ stigma.

1. In **sporophytic systems**, hindrance of pollen tube growth is localized on the surface of the stigma.
   However, an exception is **gametophytic self-incompatibility** in several grass species such as rye and ryegrass, where pollen tube growth is inhibited on the surface of the stigma.

2. In **gametophytic systems**, growth is impeded in the style.

Pollen tube growth inhibition



A feature that distinguishes sporophytic from gametophytic self-incompatibility is the location of pollen tube growth inhibition.

When there are multiple **S** alleles in sporophytic systems, genetic segregation ratios become complex. The presence of a dominant allele in the pollen-producing plant conditions incompatibility if the female carries that same dominant allele. Assume a diploid species and dominance: $\mathbf{S_1} > \mathbf{S_2} > \mathbf{S_3} > \mathbf{S_4}$.

Female Genotype	Male Genotype	Pollen Genotypes	Offspring	Explanation
S_1S_2	S_1S_2	S_1 and S_2	No	Matching alleles so incompatible
S_1S_2	S_2S_3	S_2 and S_3	No	$\mathbf{S_2}$ is imprinted on $\mathbf{S_3}$ pollen grains.
S_1S_3	S_2S_3	S_2 and S_3	Yes 1/4 $\mathbf{S_1S_2}$: 1/4 $\mathbf{S_1S_3}$: 1/4 $\mathbf{S_2S_3}$: 1/4 $\mathbf{S_3S_3}$	$\mathbf{S_2}$ is imprinted on $\mathbf{S_3}$ pollen grains.

Self-incompatibility Systems

Many crops have self-incompatibility systems. Why might these systems have evolved? Self-incompatibility is common among naturally cross-pollinated species. Self-incompatibility prevents or limits self-fertilization and promotes out-crossing. Out-crossing maintains heterozygosity and heterogeneity in a population, which often improves plant vigor and

productivity. In some species, homozygosity can severely reduce vigor, a phenomenon referred to as 'inbreeding depression.'

Table 1 Examples of crops with self-incompatibility.

Gametophytic Self-incompatibility	**Sporophytic Self-incompatibility**
Alsike clover	Sunflower (wild populations)
Red clover	Buckwheat
Tall fescue	Cacao
Potato	*Brassica* species
Rye (a two-loci system)	Cabbage
Sugar beet (a four-loci system)	Broccoli
Alfafa	Kale
Tobacco	Brussel sprouts
	Not found in monocots!

Techniques to Overcome Self-incompatibility

In order to self-pollinate or mate closely-related plants that are normally self-incompatible, plant breeders can employ various techniques that bypass self-incompatibility mechanisms. The technique used depends on the type of self-incompatibility.

Uses of Self-incompatibility Genes

Some systems have been developed or proposed for utilizing self-incompatibility genes to control pollination and produce F_1 hybrid varieties.

Gametophytic System

Cross-pollination of vegetatively propagated, self-incompatible clones. Using this approach, seed can be obtained from species that are normally self-incompatible, and thus propagated vegetatively. An example is the production of a hybrid variety of bahiagrass, Tifhi.

Sporophytic System

Bud-pollination can be used in Brassicaceae for self-pollination and thus inbreeding, resulting in inbred lines or families homozygous for an allele at the self-incompatibility (SI) locus (only one SI locus is present in, e.g., cabbage). Hybrids can subsequently be produced by planting two such lines with different fixed SI alleles side by side. Any seed produced is expected to be hybrid seed, since self-pollination within parental lines is prevented by SI. The resulting hybrid will be heterozygous for the SI locus and thus self-incompatible. However, this is not critical since the economic product does not require pollination and is vegetative, e.g., cabbage and kale.

Pseudo-Compatibility

Produce inbreds in environments that promote pseudo-compatibility and their hybrids in environments that prevent self-fertilization. Sugar beet is normally self-incompatible. However, when grown at high elevations, plants are self-compatible.

Low Elevation: Sugar beet inbreds will behave as normal self-incompatible lines at low elevations — self-pollination is genetically eliminated. All seed produced at the lower elevation will be the result of out-crossing among inbred lines, and thus will be hybrid.

High Elevation: Self-fertilized seed can be obtained via elevation-induced pseudo-self compatibility. Sufficient quantities of selfed seed can be produced for subsequent hybrid generation.

Gametophytic System

- **Pseudo-compatibility**: Expose the plant(s) to lower or higher temperatures, elevated CO_2 concentration, or electric shock to induce a pseudo-compatibility response.
- **S_f alleles**: Self-fertility (Sf) alleles, reported to be present in some species can be

transferred into a population using conventional breeding methods. The presence of an Sf allele allows self-fertilization. These alleles can be present at self-incompatibility loci, but can also be present at distinct self-fertility loci.

Sporophytic System

- **Removal of stigmatic surfaces**: Mechanical (e.g., rupturing) or chemical removal of the stigma surface eliminates the inhibiting factors [believed to be specialized proteins or enzymes, (Barrett, 1998)] that inhibit pollen tube penetration of the style. With those factors absent, pollen tube growth can proceed normally and fertilization can be achieved.
- **Bud pollination**: Pollen is placed on an immature stigma before the inhibiting factor is formed.

An example of a gametophytic system is the production of a hybrid variety of bahiagrass, Tifhi.

Male Sterility

Male sterility is due to the failure of a plant to produce functional anthers or pollen; usually, its female gametes are normal. Thus, male sterility prevents self-pollination and can be used to ensure cross-pollination without emasculation. Male sterility can have genetic causes or be induced chemically.

Female sterility, the failure of a plant to produce functional ovaries or eggs, can also occur, but is of little use to plant breeders and will not be discussed.

Fig. 3 Stamens of the Amaryllis flower. Photo by André Karwath. Licensed under CC-BY-SA 2.5 via Wikimedia Commons.

Genetic-based Male sterility

Male-sterile gene expression may be complete or partial, and may vary with environment. Breeders desire complete expression that is stable regardless of environment. The extent of sterility is measured by the percentage of viable pollen produced or percentage of seed

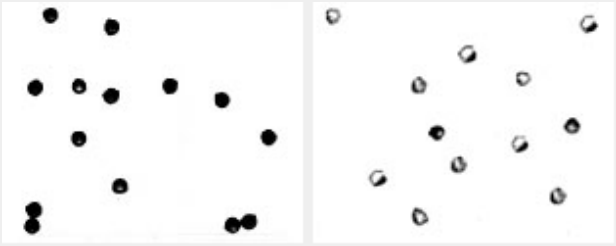
set. Male sterility is used by breeders to eliminate the necessity of emasculation to control pollination — male-sterile plants cannot self-pollinate.

Viable Pollen Measurement

The percentage of viable pollen can be estimated using a microscope. Fresh pollen from a plant or group of plants is placed on a microscope slide and viewed at low magnification, 10X.

In the field, a hand-held magnifier is often adequate for scoring the percentage of viable pollen. Pollen grains can also be stained and viewed under a microscope in the lab.

Pollen viability can also be tested in bioassays.



Soybean pollen grains. Viable pollen grains (A) appear solid. Sterile or non-viable pollen grains (B) look like miniature doughnuts, have a translucent center, or are irregularly shaped. Photo by Palmer and Kaul, 1983; used with permission.

Male-sterility is controlled by nuclear genes, the cytoplasm, or by genetic interaction between the cytoplasm and nucleus.

Controlled by the action of specific gene(s) in the nucleus. Usually, the recessive allele(s), designated **ms**, conditions inhibition of normal anthers or pollen development. Thus, the male sterility phenotype is expressed in plants homozygous for the **ms** allele.

- **ms ms** = male sterile (therefore, is functionally a female plant)
- **Ms** _ = male fertile (normal)

Maintenance of sterility genes in a population is challenging.

Uses of Genetic Male Sterility

Numerous crops have genes causing male sterility in their gene pools.

There are two major uses of genetic male sterility:

- Eliminate hand emasculation in making crosses
- Increase natural cross-pollination in populations of self-pollinated crops

These uses necessitate the transfer of the **ms** allele into the population being worked with; then the ms allele is maintained in the population through selection for sterility in each subsequent generation.

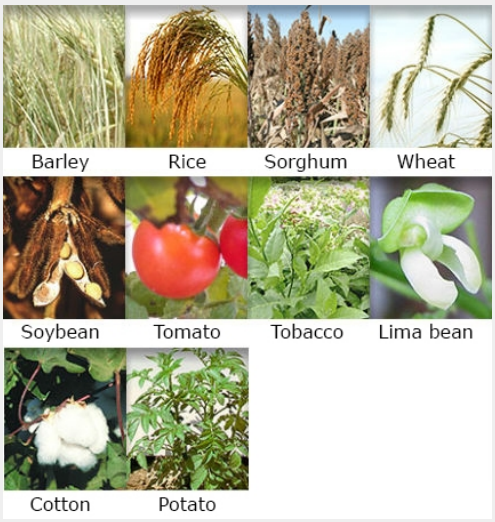


Fig. 4 Examples of crops with genes causing male sterility in their gene pool. Photos by Iowa State University. Licensed CC-BY-SA 3.0.

Because genetic male sterility is controlled by a recessive gene(s), it is not possible to get a true-breeding or homozygous ms population. However, the recessive **ms** allele can be maintained at a high frequency in the population.

Cytoplasmic Male-sterility

The genetic composition of the cytoplasm determines male sterility. The genetic makeup of the cytoplasm results from genes located in mitochondria.

Cytoplasm is inherited entirely through the female line. The cytoplasm can be

- Normal (N) — normal development of anthers and pollen = male fertile; or
- Sterile (S or CMS) — anthers or pollen are non-functional = male sterile

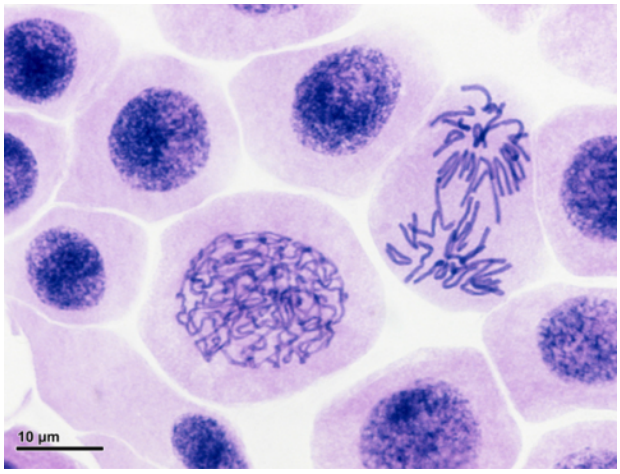


Fig. 5 Onion cells stained to show contrast between the nucleus (dark) and the cytoplasm (lighter purple). Two cells at center are undergoing mitosis. Photo licensed under CC-BY-SA 2.0 via Wikimedia Commons.

Female Line

In all organisms, the cytoplasm contains genetic material in the mitochondria. Unlike the genetic material in the nucleus, however, cytoplasmic genetic material is not subject to recombination. Cytoplasm is inherited strictly through the female parent because the male parent does not contribute cytoplasm to the zygote. Recall that the zygote is formed when the sperm nucleus and the egg nucleus fuse; since the zygote's nucleus is contained in the cytoplasm of the egg and the sperm has no cytoplasm, the zygote inherits its cytoplasm only from its female parent.

This fact is used by evolutionary biologists to trace family lines, avoiding the complexity resulting from genetic recombination and segregation that occur with nuclear genetic material.

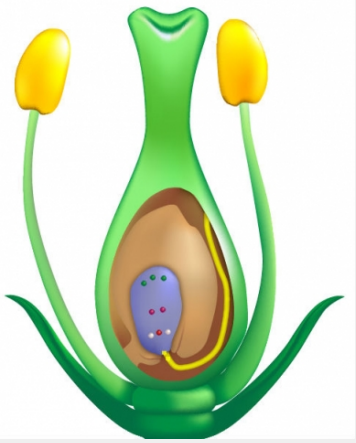


Fig. 6 Sperm cells fertilize the egg inside the cytoplasm of the female ovule.

In some systems, the expression of the male sterility phenotype depends on the interaction between cytoplasmic and nuclear genes. In these systems, a plant having sterile cytoplasm, but a nuclear dominant fertility-restorer gene (**Rf** _), will express a fertile phenotype. The particular combination of cytoplasm and nuclear genes determines the phenotype.

Table 2 Phenotype of cytoplasm and nucleus genotype combinations.

Cytoplasm Type	**Nucleus Genotype**	
	Rf_	**rf rf**
CMS	Male Fertile	Male Sterile
N	Male Fertile	Male Fertile

Plant breeders must pay close attention to both the cytoplasm type and nuclear genes of the parents used in crosses to generate and maintain cytoplasmic male sterility.

Sorghum Example

Several crops have cytoplasmic male sterility types available, including sunflower, millet, wheat, maize, and sorghum.

Let's examine the maintenance and use of cytoplasmic male sterility in sorghum. This system involves interaction between the cytoplasm and a nuclear gene. There are two types of cytoplasm possible: S = sterile; N = normal. There are two possible alleles at the restorer locus in the nucleus, **Rf** (dominant) and **rf** (recessive). In this example, we'll look at three types of inbreds.

Table 3 Genotypes and phenotypes of sorghum inbred line types.

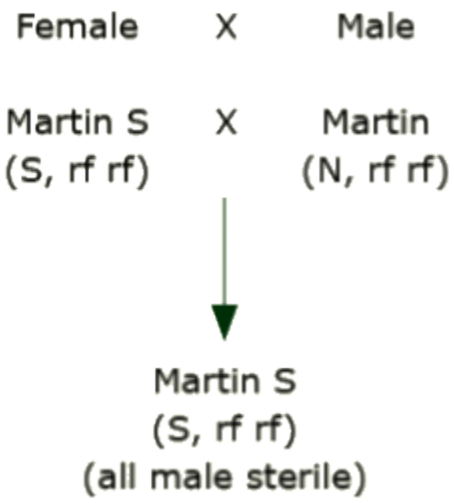
Inbred Type	**Cytoplasm Type**	**Nuclear Genotype** †	**Male Phenotype**
A line	S	rf rf	Sterile
B line	N	rf rf	Fertile (normal)
R line	N, or	Rf Rf	Fertile (normal)
	S	Rf Rf	Fertile (normal)

† **Although we use rf and Rf in this example, sorghum breeders indicate the dominant and recessive nuclear fertility restorer alleles with different symbols: msc = rf and Msc = Rf.**

How are these inbred (homozygous) lines maintained?

A line

Assume that the cultivar 'Martin' has two versions, one with normal cytoplasm and one that differs only in that it has a sterile type cytoplasm ('Martin S'). The normal version is a B inbred type and Martin S is an A type. Mate the sterile version as the female with the normal, fertile version.



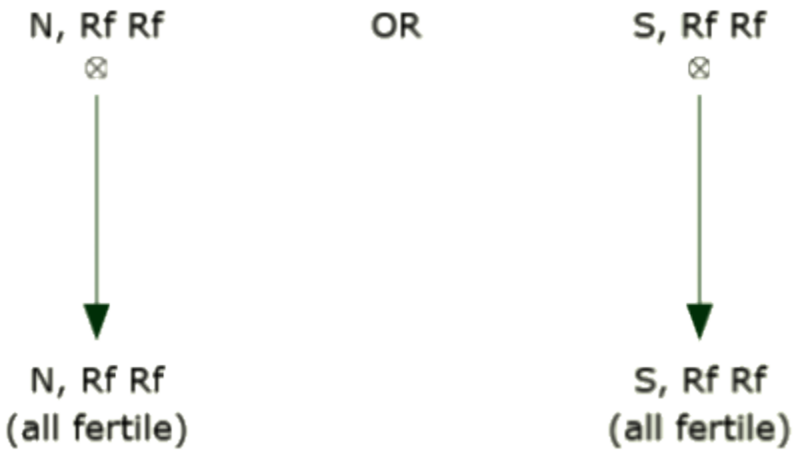
Notice that there is no blending or segregation of the cytoplasm. The F1 always has the cytoplasmic type of its female parent.

B line

B and R lines can be maintained simply by self-pollination. This is usually accomplished by growing each line in an isolated block and allowing the plants to intermate. Since the line is homozygous and fertile, the seed produced will also be homozygous and fertile.

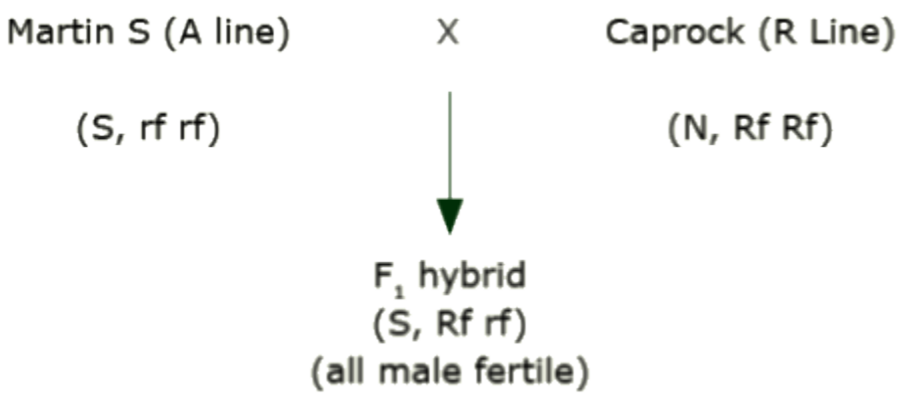
R line

Like the B line, the R line can be maintained simply by self-pollination. This is usually accomplished by growing each line in an isolated block and allowing the plants to intermate. Since the line is homozygous and fertile, the seed produced will also be homozygous and fertile.



How is single-cross hybrid seed produced using cytoplasmic male sterility?

Cross a male sterile line (e.g., A line) by a fertile line (e.g., R line). 'Caprock' is an R line.



Engineered Genetic Male Sterility

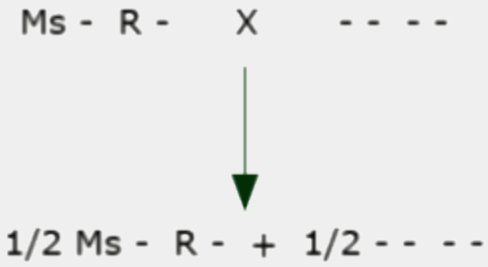
Different systems of male sterility and fertility restoration are being worked on by various companies. All of these involve molecular techniques (genetic engineering) and the development of transgenic plants.

One of the first to be utilized was the development of male sterile plant by transforming plant cells with a bacterial gene. This male sterility is dominant, since when the gene is present the plants are male sterile. The normal state for the plants is that the male sterile gene is not present and that gives normal, male fertility. As with the naturally occurring genetic male sterility we discussed earlier, it is impossible to maintain a population of completely male sterile plants as we can in cytoplasmic male sterility; so to identify the male sterile plants from the male fertile ones and to be able to eliminate male fertile plants, an herbicide-resistance gene has been linked to the male sterility gene. This resistance is also dominant, since when present it confers resistance and when absent (the normal state of the untransformed plants) plants are susceptible to the herbicide.

System: Linking herbicide-resistant gene to the male sterility gene

A plant is transformed to contain linked male sterility and herbicide resistance genes Ms R (The transformed plant will have only one homologue containing both the Ms and R genes, thus will neither be homozygous nor heterozygous, but rather what we call hemizygous for the two genes.)

We now must incorporate this linked pair into an elite inbred line by backcrossing. (The symbol "-" indicates that there is no allele for this gene present.)



The elite inbred will be developed and then maintained by crossing the male sterile plants with the normal inbred to give 50% sterile and 50% fertile plants.

Discussion for Further Thought

For the diploid, self-pollinated species tomato (*Solanum lycopersicum*), discuss from a genetic perspective the strengths and weaknesses of different hybridization systems that use either hand emasculation, male sterility, or genetic engineering, and come to a consensus, on which one you prefer.

Chemically-induced Genetic Male Sterility

In species that lack genetic male sterility or cytoplasmic male sterility, or in species in which these are difficult to work with, breeders can use chemicals to induce male sterility. These chemicals have at one time or another been called by the following names:

- Gametocides
- Pollen suppressants
- Hybridizing agents

These inhibit the production of viable pollen or prevent pollen shed, but do not damage the pistillate flower or interfere with seed development. There are both advantages and disadvantages to chemically-induced male sterility.

Advantages

- No need to develop and maintain cytoplasmic or genetic male sterility systems
- Applied to normally self-pollinated species, the chemicals prevent self-pollination and facilitate cross-pollination without the necessity of hand emasculation. This approach has been used in cotton, maize, sorghum, and various vegetable crops. It was also tested for hybrid wheat production.

Disadvantages

- Incomplete pollen sterility can occur, resulting in some selfing. Several factors can account for incomplete sterility.
    - differential genotypic reactions to the chemical agents

    - timing and dosage of application is critical and varies with genotype
    - environmental effects on chemical and interaction with genotype
    - long periods of flowering cause difficulties in maintaining optimum dosage in plants
- Does not provide a means of pollen transfer to produce hybrid seed, and so additional methods must still be employed.

Hybrid Wheat

Since wheat has perfect flowers and is normally self-pollinating, the production of hybrid varieties was impractical without the use of male gametocides. Gametocides were applied to the female plants prior to flowering to kill their pollen and prevent their self-pollination; the treatment was not applied to adjacent rows containing the intended male parent. Thus, seed produced by the female plants resulted from cross-pollination with the desired males. Although hybrid seed could be produced in this manner, the approach has been largely abandoned by most commercial seed companies—farmers did not buy the hybrid seed because the hybrids did not yield sufficiently better than conventional wheat to justify the additional seed costs.

Sex Inheritance

Description

The ultimate biological mechanism to prevent self-pollination, and thus inbreeding, is to have distinct male and female flowers on different plants. In such “dioecious” plant species, which are common in the plant kingdom, male and female plants occur usually at a 1:1 ratio. The decision on whether a plant becomes male or female is genetically determined, and can in some species be influenced by environmental factors. As a consequence, development of inbred lines can be accomplished either by crosses of male and female sister plants, or by employing environmental factors to induce plants with both sexes to allow self-pollination. In addition to dioecious and hermaphrodite plant species, there are various intermediate variants realized in plants (Table 4).

Table 4 Sex systems in flowering plants. Data from Charlesworth 2002, Heredity 88, 94-101.

Plant Term	Definition of Plant Term
Sexually monomorphic	
Hermaphrodite	Flowers have both male and female organs (Tomato, potato, sugarcane)
Monoecious	Separate sex flowers on the same plant (Cassava, maize, banana)
Gynomonoecious	Individuals have both female and hermaphrodite flowers
Andromonoecious	Individuals have both male and hermaphrodite flowers
Sexually polymorphic	
Dioecious	Male and female plants (spinach, asparagus)
Gynodioecious	Individual either female or hermaphrodite (papaya)
Androdioecious	Individuals either male or hermaphrodite

Genetics of Sex Inheritance

Distinct male and female flowers can be of practical importance, if present on the same plant. Best example is maize, where spatial separation of female and male flowers facilitates both elimination of male flowers to produce "female plants" for hybrid seed production, and self-pollination for inbred development. However, the majority of plants carry either flowers containing both male and female organs, or male and female flowers on distinct plants. In the latter case, pollination will be mediated by wind or insects. Female plants will set seed only if a male plant is located in the vicinity.

There are two genetic mechanisms of sex inheritance:

- **Sex chromosomes** — Like in humans, the sex of an individual is determined by specific chromosomes. In the case of humans, carriers of two X-chromosomes are females, whereas males have one copy each of the X- and Y-chromosomes. Since X- and Y-chromosomes differ cytologically, their pairing in meiosis might be incomplete.
- **Autosomal inheritance** — Genes affecting sex inheritance are located on "regular" chromosomes (autosomes), forming bivalents in meiosis. Depending on the plant species, sex determination can be due to one or a limited number of genes (mono-, or oligogenic inheritance).

Sex Inheritance Systems

Cucumber Mode of inheritance: Autosomal, 3 genes

Asparagus Mode of inheritance: Sex chromosomes

Spinach Mode of inheritance: Autosomal, 1 gene. (Image licensed CC-BY-SA 3.0 via Wikimedia Commons.)

Hops Mode of inheritance: Sex chromosomes. (Image licensed CC-BY-SA 3.0 via Wikimedia Commons.)

Application and Challenges

Application

Like self-incompatibility or male sterility, sex inheritance can be used for controlled crosses between any pair of genotypes in hybrid seed production schemes. One genotype would be used as female and seed parent, the other parent as pollen donor.

Challenges in using sex inheritance for hybrid seed production

As plants with only one sex cannot be sexually maintained, it is either required to use environmental or chemical factors to induce the other sex to allow self-pollination (if available),

or to develop male and female sister lines for each hybrid parent, comparable to male sterile and maintainer lines when using male sterility. This is described in more detail in the following cucumber and asparagus examples.

Cucumber

Cucumber (*Cucumis sativus*) has a wide range of floral types. Staminate (male), pistillate (female), and hermaphrodite flowers can occur in different arrangements. Generally, embryonic flower buds possess both staminate and pistillate initials, and thus the potential to develop into any of the above-mentioned flower types. The cucumber phenotype with regard to floral types depends on autosomal genes and their interaction with environmental factors.

Sex inheritance is controlled by a minimum of three major loci:

- **m^+, m** — controls the tendency to form hermaphrodite versus male or female flowers. **mm** homozygotes develop hermaphrodite flowers.
- **F^+, F** — controls the female tendency, with **F** allele being dominant and favoring female flowers. This locus is subject to strong environmental influence. Among environmental factors, in particular photoperiodic conditions and temperature affect flower formation.
- **a^+, a** — Homozygotes for **a** allele intensify male tendency. The effects of a are dependent on the allelic composition at the 'F' locus. Male tendency is only pronounced in 'F^+F^+' homozygotes.

Although other loci such as 'de' and 'cp' have been shown to affect sex types in cucumber, the main sex types are determined by the above mentioned loci **m**, **F**, and **a** (Table 5).

Table 5. Phenotypes and Genotypes of basic sex types in cucumber.Note: "-/-" means that any allele can be present at this locus

Genotype, locus			
Phenotype	**m**	**F**	**a**
Androecious (male)	-/-	F/F	a/a
Monoecious	m^+/m^+	F/F	-/-
Hermaphroditic	m/m	F/F	-/-
Gynoecious (female)	m^+/m^+	F/F	-/-

CHEMICAL REGULATION OF SEX EXPRESSION

Phytohormones can be applied to alter flower phenotypes. Whereas auxin and ethylene promote female flowers, gibberellin promotes male flowers. Moreover, silver nitrate and silver thiosulfate induce male flowers even in strongly gynoecious genotypes. These treatments have become invaluable in cucumber breeding, as they allow efficient development of gynoecious inbreds to be used as female hybrid parents, and even the production of hybrids from crosses of two gynoecious lines.

Asparagus

Sex inheritance in *Asparagus officinalis*, is similar to humans, based on sex chromosomes. XX chromosome carriers are female, XY carriers are male. Males and females occur at about equal frequencies in natural populations. Males have been shown to be more productive. For this reasons varieties with only male plants are preferred. At a low frequency and for unknown reasons from a genetic perspective, andromonoecious XY plants occur. These plants can be self pollinated, and will result in a 1:2:1 ratio with regard to the sex chromosomes in resulting offspring. One quarter of those will thus carry two copies of the "male" Y chromosome, and are thus called supermales. As Asparagus can be vegetatively propagated, those YY plants are potential variety candidates. Alternatively, they can be used as male parent in hybrid breeding schemes to be crossed with female XX genotypes, and result in 100% male XY offspring. In contrast, if regular XY males would be crossed to XX females, the offspring would segregate 1:1 into XY and XX plants, and XX plants would need to be eliminated by producers.

Fig. 7 Asparagus officinalis plants. Photo by Rasbak. Licensed under CC-SA 3.0 via Wikimedia Commons.

Controlling Hybridization

Knowledge of the breeding system of a crop species is essential to take advantage of the types of gene action that give the most useful cultivars. Self-incompatibility systems are important in many natural species for forcing outcrossing and thus maintaining vigor through

heterozygosity. As we have seen, this system can be adapted to help produce F_1 hybrid cultivars in domesticated species containing self-incompatibility loci. Similarly, sex inheritance can be employed in controlled F_1 hybrid seed production schemes.

Male sterility is a rather unimportant method of enforcing outcrossing in natural plant species. It has, however, become such an important tool in the production of hybrid cultivars that it is utilized in many species of both cross- and self-pollinated crops.

References

Barrett, Spencer C. H. 1998. The evolution of mating strategies in flowering plants. Trends in Plant Science 3(9):335-341.

Charlesworth, D. 2002. Plant sex determination and sex chromosomes. Heredity 88:94-101.

Ellison, J.H. 1986. Asparagus breeding, Chapter 14, pp. 521-569, in M.J. Bassett (editor). Breeding Vegetable Crops. AVI Publ. Co., Westport, CT.

Goldberg, R.B., T.P. Beals, and P.M. Sanders. 1993. Anther development: basic principles and practical applications. Plant Cell 5:1217-1229.

Lower, R.L., and M.D. Edwards. 1986. Cucumber breeding, Chapter 5, pp. 173-207, in M.J. Bassett (editor). Breeding Vegetable Crops. AVI Publ. Co., Westport, CT.

Palmer, R. G., and M. L. H. Kaul. 1983. Genetics, cytology, and linkage studies of a desynaptic soybean mutant. J. Heredity 74:260-264.

Pierce, B. A. 2008. Genetics: A Conceptual Approach. 3rd edition. W.H. Freeman, New York.

How to cite this chapter: Muenchrath, D., A. Campbell, L. Merrick, T. Lübberstedt, and S. Fei. 2023. Controlled Hybridization: Self-incompatibility, Male-sterility, and Sex-inheritance. In W. P. Suza, & K. R. Lamkey (Eds.), *Crop Genetics*. Iowa State University Digital Press. DOI: 10.31274/isudp.2023.130

Gene Segregation and Genetic Recombination

Deborah Muenchrath; Arden Campbell; Laura Merrick; Thomas Lübberstedt; and Shui-Zhang Fei

Introduction

Plant breeders take advantage of the variation that occurs within a population to develop improved cultivars. Ordinarily, the goal of the plant breeder is to combine the favorable characteristics of one plant or cultivar with the desirable traits of another plant or cultivar to obtain a new combination that has the best of both. Understanding the genetics of desired, as well as undesirable, characteristics enhances the efficiency of the plant improvement process.

Learning Objectives

- Understand the molecular basis of genes and chromosomes.
- Understand the basic principles of transcription and translation.
- Understand Mendelian mechanisms and patterns of inheritance.
- Be able to differentiate among different types of gene action.
- Determine genotypic and phenotypic consequences of independently inherited genes through generations of self-pollination.
- Know how epistasis occurs through interaction of genes and alteration of expected phenotypic ratios.

Overview of Genetics

The Science of Genetics

Genetics is one of the principal sciences that underlie plant breeding. Genetics is the study of heredity, **genes**, chromosomes, and variation in biological organisms. The science of genetics is often divided into four major subdisciplines:

- Transmission genetics (also called classical or Mendelian genetics)
- Quantitative genetics
- Population genetics, and
- Molecular genetics

Transmission genetics deals with how genes and genetic traits are transmitted from generation to generation and how genes recombine. The foundation of modern genetics is recognized to have occurred in the mid-1800s when Gregor Mendel analyzed the results of crosses he made among garden pea plants. Mendel concluded that inherited characteristics (now called traits or phenotypes) are determined by factors (now known as genes) that he observed. He also realized that each organism contained two copies of each "factor" (gene), one inherited from its mother and one from its father. Mendel discovered the principles of heredity when he noticed how inherited traits (e.g., seed shape round vs. wrinkled; pod color yellow vs. green; flower position axial vs. terminal; or plant height tall vs. short) are passed from parents to offspring. Transmission (Mendelian) genetics is the focus of this module.

Genetic Subdisciplines

Quantitative genetics focuses on the study of inheritance when phenotypes exhibit continuous variation or distribution. In particular, it considers the effects of many genes that could be simultaneously influencing such traits, as well as the relative contributions of the environment and the interaction between genotype and environment. Quantitative genetics is the focus of the module on Inheritance of Quantitative Traits.

Population genetics entails a study of heredity in groups of individuals for traits that are usually determined by one or only a few genes. It deals with gene distribution and genetic diversity within and among populations and subpopulations. Population genetics includes assessment and prediction of response to selection. It describes relationships between allele and genotype frequencies due to four main evolutionary forces: natural selection, genetic drift, mutation, and gene flow. Population genetics is the focus of the module on Inbreeding and Heterosis.

Molecular genetics is concerned with the molecular structure and function of genes. It includes the study of DNA structure and replication and deals with gene expression and regulation.

Gene Structure

Genes and Chromosomes

To understand inheritance, it is essential to understand gene structure and action. Let's review key terminologies and principles. For a more in-depth review, please refer to biology or genetic textbooks, for example, *From Genes to Genomes (Hartwell et al. 2011), Genetics: A Conceptual Approach (Pierce 2012), or iGenetics: A Molecular Approach (Russell 2010).*

Genes are encoded with DNA. Most of the DNA in plants is located in the nucleus of cells and arranged in groups of genes along multiple, linearly-shaped, chromosomes. Nuclear DNA is subject to Mendelian inheritance, which will be discussed later in this module. In addition to its occurrence in chromosomes in the nucleus, DNA is also located in organelles present in the cytoplasm of plant cells.

FYI: Cytoplasmic DNA

In plants, DNA is not just present in the nucleus of cells. It is also located in other membrane-bound, specialized subunits known as organelles that are found within the cytoplasm, or cell fluid. Two plant cell organelles that contain DNA are chloroplasts (which are plastids or organelles that carry pigments-specifically green chlorophyll) and mitochondria (singular, mitochondrion; organelles that break down complex carbohydrates and sugars into usable forms, and thus supply energy for the plant).

The non-nuclear, organellar DNA located in plants follows cytoplasmic inheritance and is not subject to Mendelian inheritance. Cytoplasmic inheritance is also known as extrachromosomal or extranuclear inheritance, and is of significance in certain types of male sterility where the genes for those traits are present in the mitochondria, not in nuclear chromosomes.

Molecular Basis of Chromosomes

Chromosome – Each chromosome contains a single DNA molecule (Fig. 2).

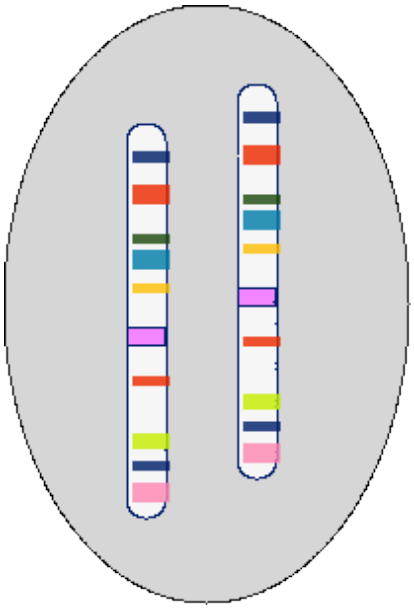


Fig. 1 Nucleus with one pair of homologous chromosomes.

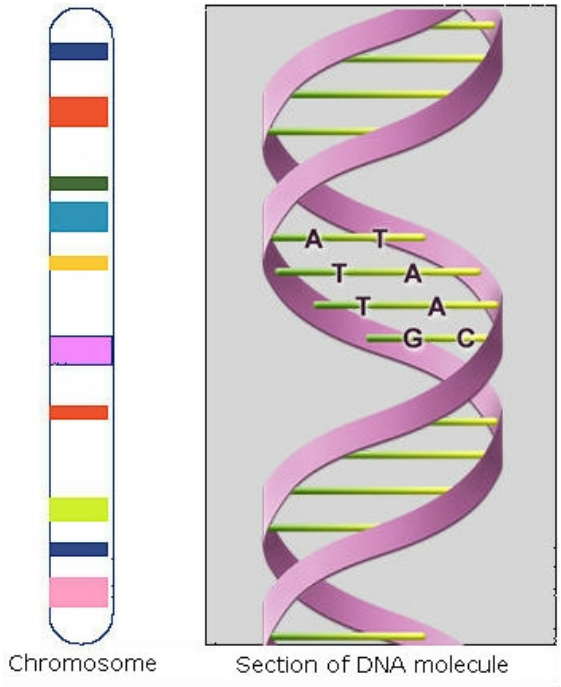


Fig. 2 DNA molecule forms a double helix.

DNA

DNA (**deoxyribonucleic acid**) is composed of two chains of polynucleotides. Polynucleotides are also called nucleic acids, and consist of linear polymers that are macromolecules formed by the chemical joining of many identical or similar units called nucleotides. Every nucleotide in each chain consists of a nitrogen-containing base, deoxyribose (a sugar), and a phosphate group. Nucleotides within each chain are held together by sugar-phosphate (phospho-diester) bonds (Fig. 3).

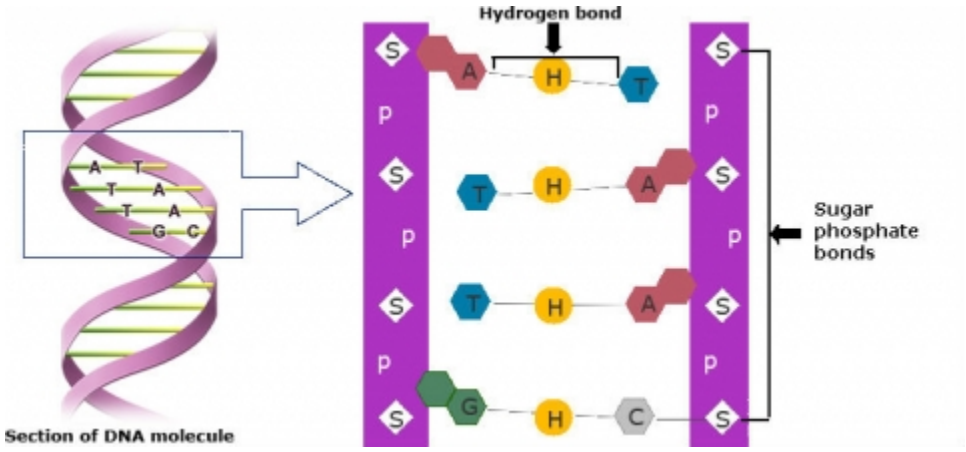


Fig. 3 Bonds holding together nucleotides in DNA chains.

Nitrogen-containing bases are purines (adenine, A, and guanine, G) and pyrimidines (cytosine, C, and thymine, T). Pairing occurs between one purine and one pyrimidine and is specific. Sequences of consecutive nucleotides constitute genes (Fig. 4).

C always pairs with G

T always pairs with A

DNA replication is semiconservative.

The process of DNA replication is not yet fully understood. Basically, there are three steps.

1. Two strands of DNA unwind and pull apart.
2. Free (unbound) nucleotides bind to complementary bases on an original strand of DNA.
3. One newly formed strand and a template DNA strand re-coil to form a double helix.

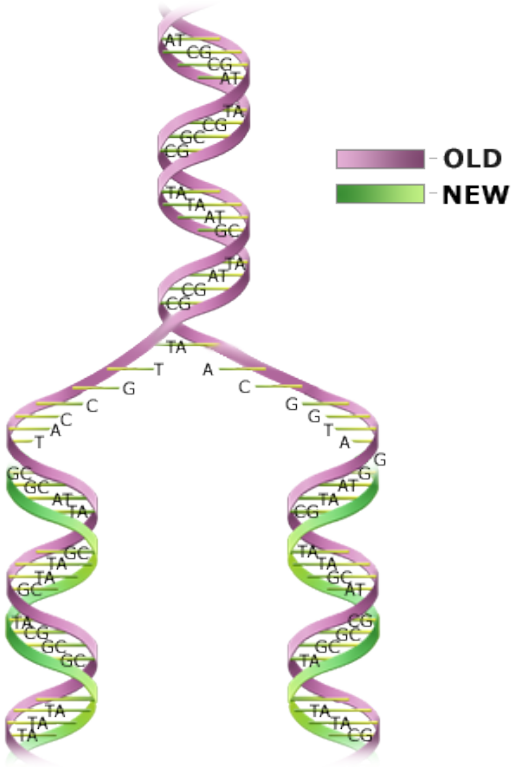


Fig. 4 Semiconservative replication of DNA.

This process is semiconservative because each resulting double-stranded DNA molecule is composed of a newly synthesized strand and a template strand (Fig. 4). Since one strand of

each DNA molecule is an original strand, there is less probability of error occurring during replication.

Genes

Many genes are present in each chromosome. Each specific gene occurs at a defined point on a chromosome, the gene **locus**, on each of the two **homologous chromosomes**. More than one form of a particular gene, **alleles**, may occupy the same locus on homologous chromosomes.

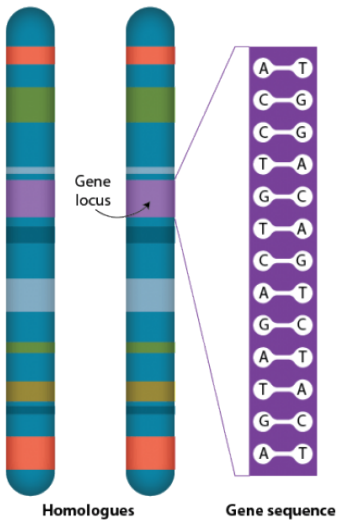


Fig. 5 Homologues showing loci of several genes.

Alleles

Alleles are variants that differ slightly in their DNA sequence. Diploid plant species have two sets of chromosomes, each of which can possess a different allele for a particular gene. For example, a gene for seed color might have the two alleles, **A** and **a**. Allele **A** causes one phenotype (e.g., brown seed color) and allele **a** causes a different phenotype (e.g., white seed color). For that gene, the genotype could be either **AA**, **Aa**, or **aa**.

If one allele at a locus on a homologous chromosome partially or completely masks the expression of the other in influencing the phenotype, the allele that is expressed is termed dominant and the allele that is masked is termed recessive. By convention, we often write the dominant form with an uppercase letter, and the recessive form in lowercase. In the example above for seed color, allele **A** is the dominant allele. If the **A** allele is completely dominant to the **a** allele, individuals with either the **AA** or **Aa** genotypes would have the brown seed color phenotype, while **aa** individuals would have white seeds.

An individual is heterozygous (**Aa**) when two different alleles are present at a locus and is homozygous—in this example, either homozygous dominant (**AA**) or homozygous recessive (**aa**)—when the same alleles are present on both chromosomes. Alleles at a locus can interact in several ways that are revealed by their phenotype, whether **heterozygous or homozygous**.

FYI: Homozygosity and Heterozygosity

For a given locus, an individual with a genotype of either **AA** or **aa** is homozygous for that gene and is known as a **homozygote**; the status of the gene is referred to as **homozygosity**. An individual with the genotype **Aa** is heterozygous for that gene and is called a **heterozygote**; the status is known as **heterozygosity**. In the case of polyploid individuals, those with the genotypes **AAAA** (tetraploid) or **aaa** (triploid) would be examples of homozygotes and those with genotypes of **AAaa** (tetraploid) or **AAaaaa** (hexaploid) would be examples of heterozygotes.

The terms homozygous and heterozygous are used to describe the status of single genes or all gene loci within an individual, not within a population. There may be many different alleles of a gene present in a population of individuals, but for each diploid individual there are only two alleles per gene. For each individual, there is one allele from each parent and each allele per gene is present at corresponding loci on homologous chromosomes.

With regard to populations, a **homogeneous population** would be one in which all individuals in the population would have the same genotype and possess the same alleles for one or more genes. In contrast, a **heterogeneous population** would be characterized by differing alleles at one or more loci. Note that a cross between two homozygous parents produces progeny that are homogeneous because all of the individual offspring are genetically identical. However, the offspring would be heterozygous for all loci for which different alleles occurred in the two parents.

Gene Expression, Translation, and Transcription

DNA, Protein, and Other Gene Products

In order to have a better understanding of the concept of gene that will be the focus of this and the following lesson on linkage, it is critical to understand the chemical nature of **DNA** . Let's review the pathways by which the genetic information in DNA is transferred from one DNA molecule to another (the process termed DNA replication) and from DNA to **ribonucleic acid** (**RNA**) molecules (called transcription), and then transferred from RNA to a protein (termed translation) by a code that specifies the amino acid sequence of the protein (see Fig. 6).

A **gene** is a stretch of DNA along a **chromosome** consisting of sequences of consecutive **nucleotides**. Recall that genetic information in DNA is coded in the sequence of four nucleotides that are abbreviated by the type of nitrogen-containing base that each contains—the purines A and G and the pyrimidines C and T. Through DNA replication, genetic information of an individual is transmitted from cell to cell during development and from generation to generation during reproduction.

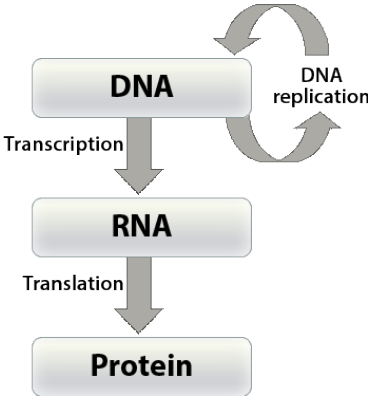


Fig. 6 Schematic of genetic information pathways from DNA to RNA protein.

DNA Structure

Examine the following for a better understanding of the chemical structure of the nucleotides that comprise the basic building blocks of DNA and the process of DNA replication:

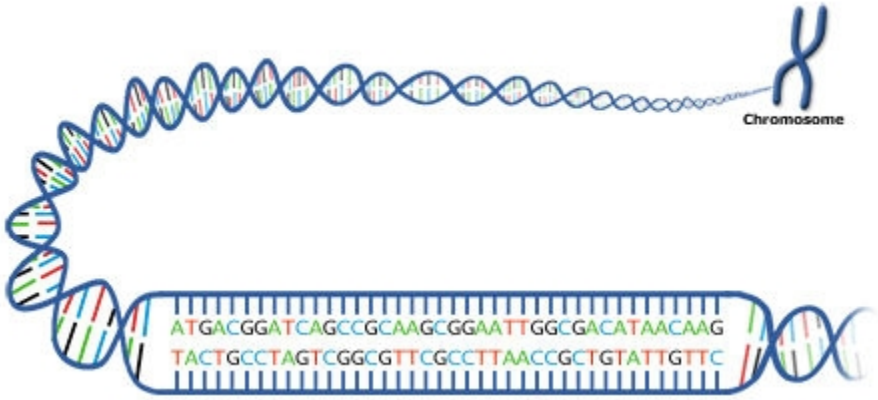


Fig. 7 Sequence of bases (A, G, C, T) within a gene carry the instructions for assembling a protein. Illustration from NIH-NHGRI, 2011

Review the chemical structure of DNA and what occurs during the process of DNA replication. DNA replication occurs within the synthesis phase of the cell cycle.

Four types of chemical bases—A, G, C, T—in gene sequences carry the instructions for assembling a protein (Fig.8). The base pairs are bonded together by H-bonds to form the "rungs of a DNA ladder" (Fig. 8).

Fig. 8 DNA molecules consist of two strands bound together by hydrogen bonds. Illustration from NIH-NHGRI, 2011

Nucleotides

Nucleotides are the basic building blocks of nucleic acids such as DNA and RNA, which are polymers made of long chains of nucleotides. DNA is double-stranded and RNA is single-stranded (Fig. 9). Note that in RNA, the chemical base **uracil** (**U**) replaces **thymine** (**T**).

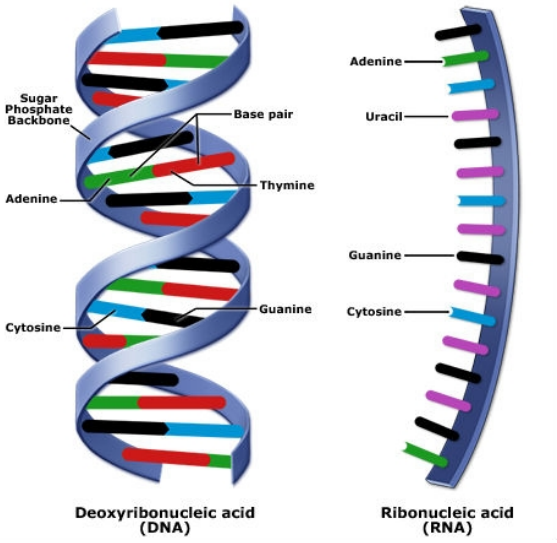


Fig. 9 Nucleic acids such as DNA and RNA are macromolecules whose function in cells is to store and express genetic information. Illustration from NIH-NHGRI, 2011.

Genes generally express their effect by coding for **polypeptide** chains, which are polymers consisting of ten or more **aminoacids** linked by **peptide** bonds. One or more polypeptides

make up a protein. The DNA sequence of a gene is used as the basis for producing a specific protein sequence. Proteins are the complex molecules responsible for most biological functions in the cell.

Gene Expression, RNA, Translation, and Transcription

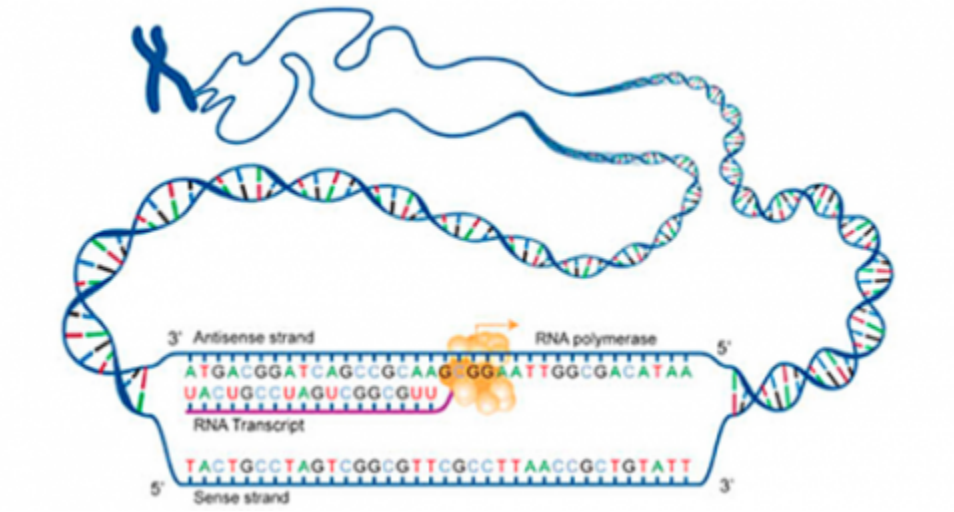


Fig. 10 Transcription is the process of making an RNA copy of a gene sequence encoded by DNA. Illustration from NIH-NHGRI, 2011.

Amino acids are the building blocks of proteins. A **protein** is composed of one or more long chains of amino acids, the sequence of which corresponds to the DNA sequence of the gene that encodes it. The process of creating proteins from the genetic code in DNA is referred to as **gene expression**. The general process of gene expression in the cells of eukaryotes such as plants involves numerous steps, which are described below.

FYI: Eukaryotes

Plants are multicellular organisms known as eukaryotes, which are organisms possessing cells that contain DNA in a nucleus and other membrane-bound, specialized subunits known as organelles that are found within the cytoplasm, or cell fluid. Two plant cell organelles that contain DNA are chloroplasts (which are plastids or organelles that carry pigments—specifically

green chlorophyll) and mitochondria (singular, mitochondrion; organelles that break down complex carbohydrates and sugars into usable forms, and thus supply energy for the plant).

The non-nuclear, organellar DNA located in plants follows cytoplasmic inheritance and is not subject to Mendelian inheritance. Cytoplasmic inheritance is also known as extrachromosomal or extranuclear inheritance, and is of significance in certain types of male sterility where the genes for those traits are present in the mitochondria, not in nuclear chromosomes.

In contrast to eukaryotes, prokaryotes such as bacteria are often unicellular and lack a cell nucleus and usually have their DNA in a single circular molecule.

Transcription is a process in which the sequence of nucleotides in one DNA strand of a gene is copied into the nucleotides of an RNA molecule. The order of nucleic acids in RNA complements those on the DNA strand from which it is transcribed. In the RNA strand, however, uracil(U), rather than thymine (T), is the base that complements adenine (A). As the RNA transcript is formed, each base in the DNA is paired with a base in an RNA nucleotide, which is progressively added to the RNA strand as it grows. Transcription occurs in the nucleus of the cell (Fig. 13).

In a procedure known as RNA processing, intervening sequences or introns are removed from the RNA transcript by splicing. Introns are a special type of so-called non-coding DNA sequences that do not code for amino acids, but are located within genes until such sequences are removed during RNA processing. (Note that aside from intron sequences, most non-coding DNA found in chromosomes is located between (not within) gene loci along the chromosome.) The regions between the introns in the fully processed RNA are called exons, the sequences that code for proteins (Fig. 10). The ends of the transcript are also modified. The fully processed RNA is referred to as mRNA (messenger RNA). mRNA is a single-stranded sequence of nucleic acid and it moves from the cell nucleus to the cytoplasm where proteins are made (Fig. 10).

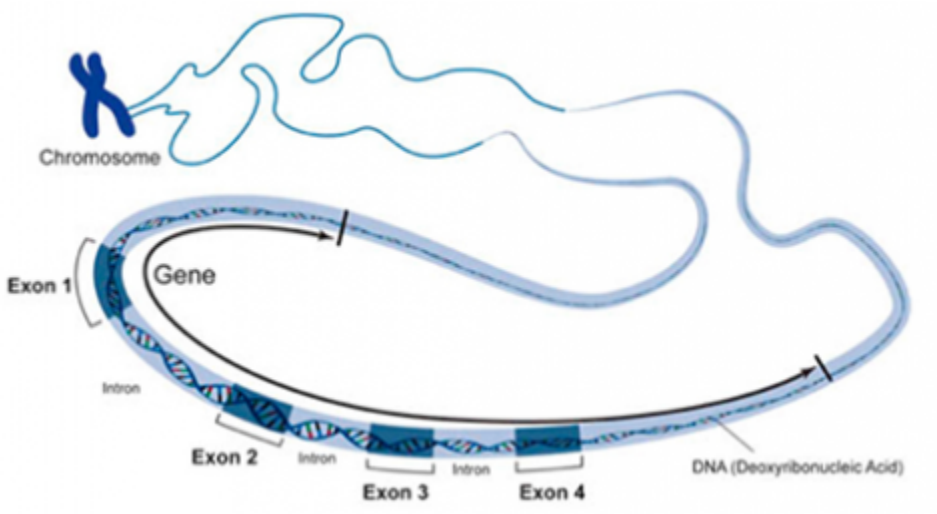


Fig. 11 During RNA processing, the non-coding DNA sequences of the gene known as introns are removed and genetic information in exons is encoded. Illustration from NIH-NHGRI, 2011.

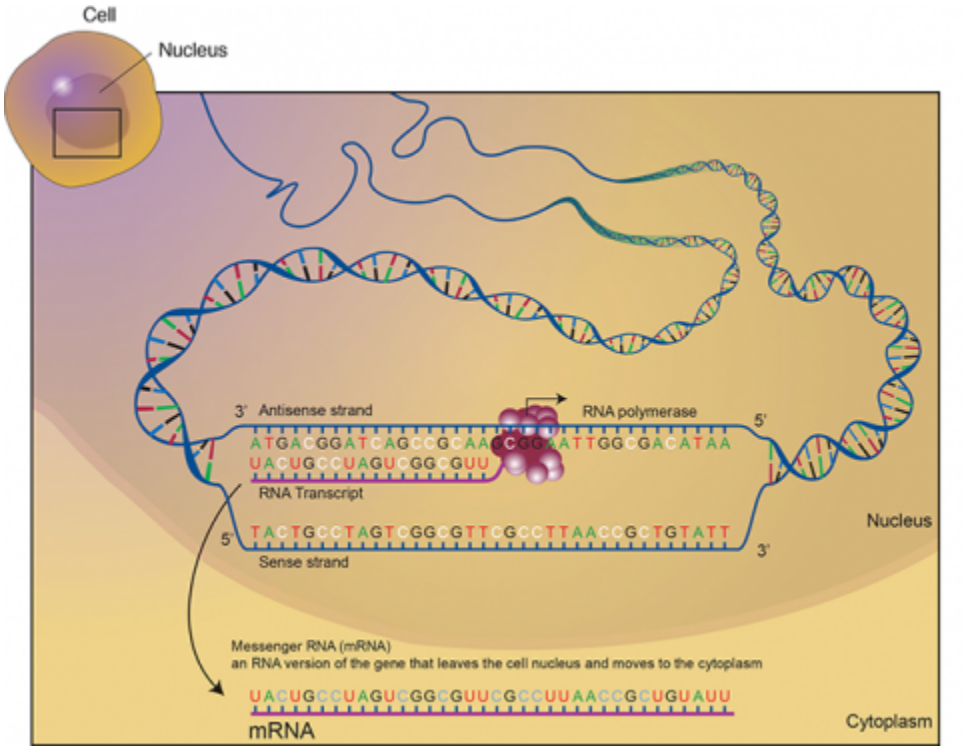


Fig. 12 Messenger RNA(mRNA) is a single-stranded RNA molecule that is complimentary to the coding regions (exons) of one of the DNA strands of the gene. Illustration from NIH-NHGRI, 2011.

Translation is the process through which mRNA directs the assembly of amino acids in the

proper sequence to synthesize the particular protein. Ribosomes in the cell cytoplasm read the base sequence of the mRNA (Fig. 10).

In the translated part of the mRNA, each adjacent group of three nucleotides constitutes a coding group or **codon**. Each codon specifies an amino acid subunit in the polypeptide chain. Adapter molecules, **tRNA** (**transfer RNA**) are complexed with the specific amino acid corresponding to the base sequence of the given mRNA. tRNA molecules bring the amino acids specified by the mRNA to the ribosomes where they are added to the growing protein chain. When the polypeptide chain is complete, it is released from the mRNA and forms a protein molecule. The order of amino acids determines the structure of the protein which affects its action.

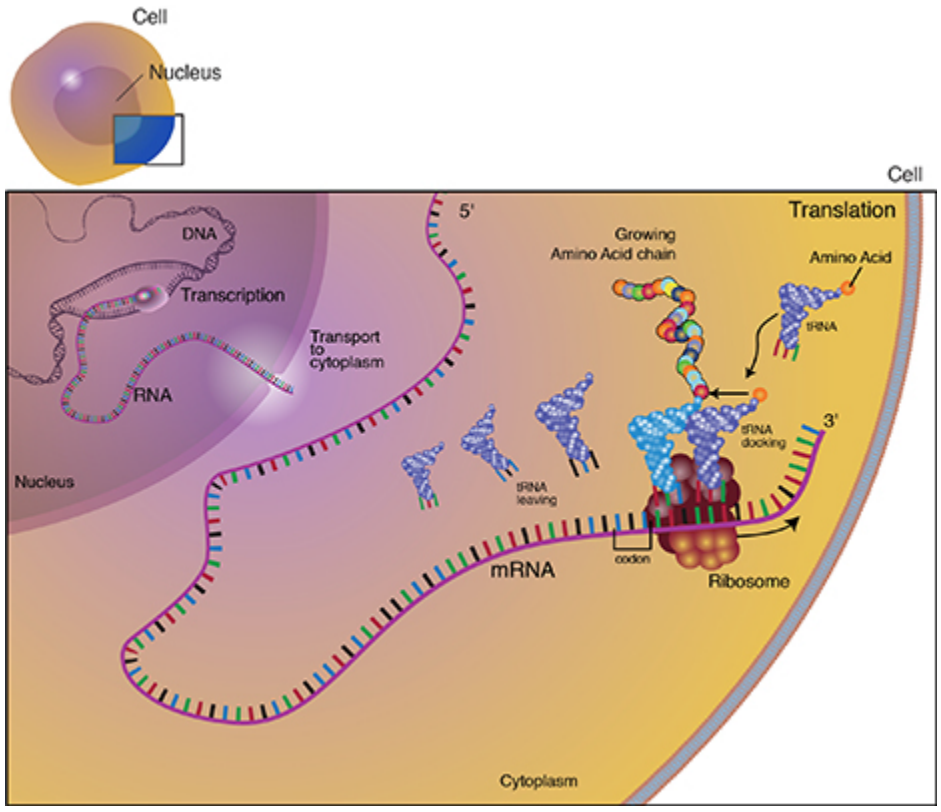


Fig. 13 Messenger RNA(mRNA) is a single-stranded RNA molecule that is complimentary to the coding regions (exons) of one of the DNA strands of the gene. Illustration from NIH-NHGRI 2011.

Basic Steps of Transcription

These are the basic steps of transcription and translation:

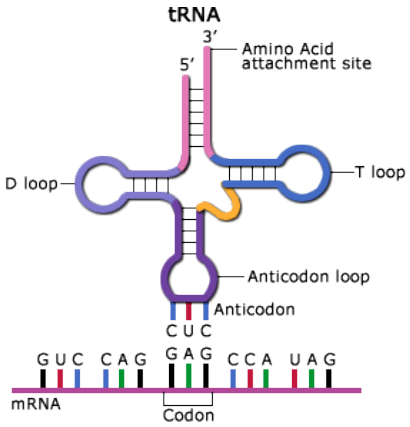


Fig. 14 Transfer RNA (tRNA). Illustration from NIH-NHGRI, 2011

1. During transcription, a region of double-stranded DNA is momentarily pushed open, separating the two strands and allowing an **enzyme** known as RNA polymerase to build a strand of mRNA corresponding to that region of DNA.
2. The tRNA anticodon attaches to the mRNA codon. The tRNA has a region called the "anticodon" that complements the codon sequence of the mRNA (Fig. 14).
3. The specific amino acid complexed with the tRNA is held in place while the tRNA-amino acid complex corresponding to the next codon moves into place. A peptide bond is formed between the adjacent amino acids, building the protein molecule.

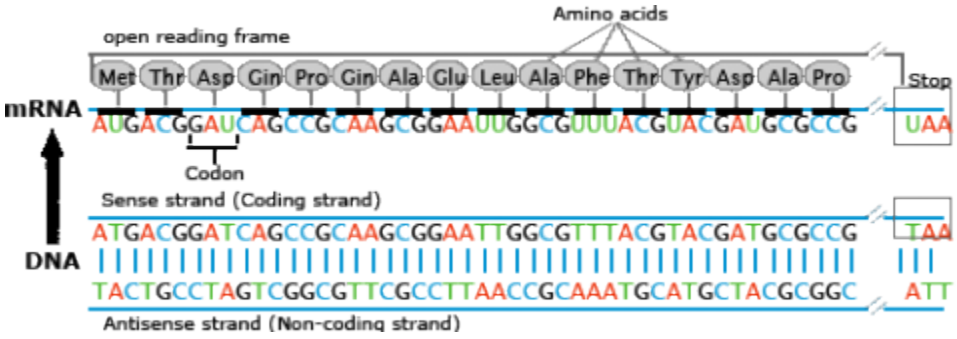


Fig. 15 An open reading frame is a fragment of DNA that when translated contains no stop codons. Adapted from NIH-NHGRI, 2011.

Inheritance and Gene Action

Mechanisms

Inheritance is based on the behavior of chromosomes and the genes that they carry. During **meiosis** and gametogenesis, homologous chromosomes separate. Each gamete receives one

(haploid) set of chromosomes. The particular chromosome of a homologous pair that is distributed to a given gamete is random. When two gametes fuse during fertilization, the zygote receives from each parent one set of chromosomes, and the alleles that they each carry. The resulting combination of alleles in the zygote determines its genotype.

Because the distribution of homologous chromosomes to gametes is random, the fusion of gametes to form the zygote may produce different genetic combinations. Thus, within a population, variation for specific traits or characters may be observed. If the variation for a given trait is due to contrasting alleles at one or more loci, rather than to responses to the environment, the variation is heritable and can be transmitted from parent to progeny. Plant breeders select plants that exhibit desirable characteristics and those plants carry the desired allele of the gene that encodes the characteristic of interest.

Each gene or combination of genes and alleles, as influenced by the environment, determines the phenotype or observed expression of the particular trait. An individual's allelic composition at corresponding loci on homologous chromosomes confers the expression of that gene. Alleles at corresponding loci interact. One allele may mask the presence of the other allele(s).

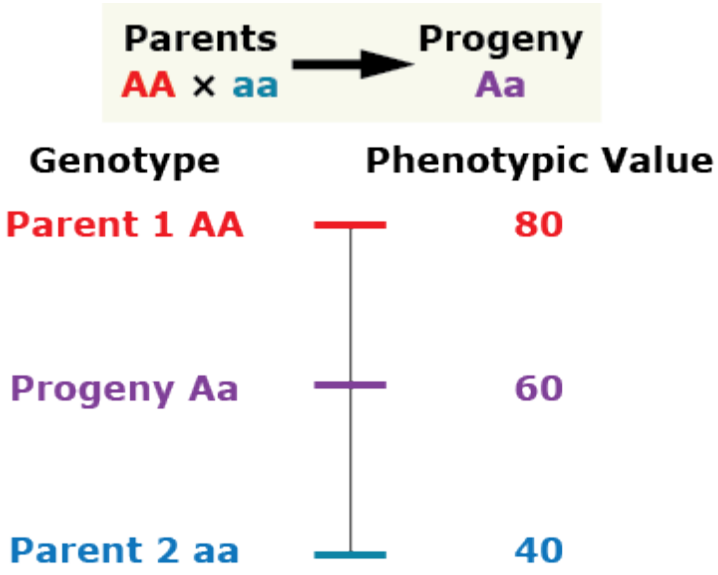
Alleles at a locus can interact in different ways, including no dominance (also referred to as additive gene action), partial dominance, complete dominance, and over-dominance.

Gene Action

There are several general types of gene action. The type of gene action and the alleles present for a given gene affect the phenotype. Let's consider the gene action as indicated by the phenotype of a diploid individual heterozygous at the given single locus compared to the phenotype of its parents.

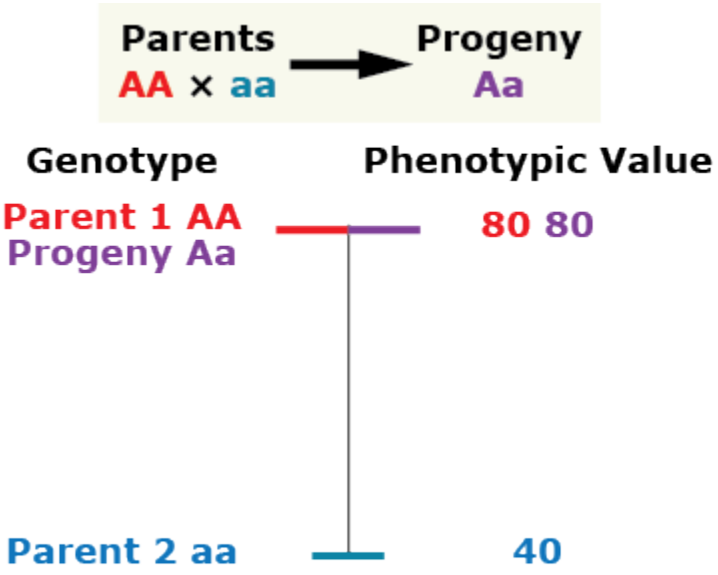
Addictive gene action (no dominance)

The progeny's phenotypic value is at the midpoint between both parents.



Complete dominance

The phenotype of the heterozygous progeny equals the phenotype of the homozygous dominant parent.



Partial (incomplete) dominance

The heterozygous progeny has a phenotypic value greater than that of the **mid-parent value** (**MPV**), but less than that of the homozygous dominant parent.

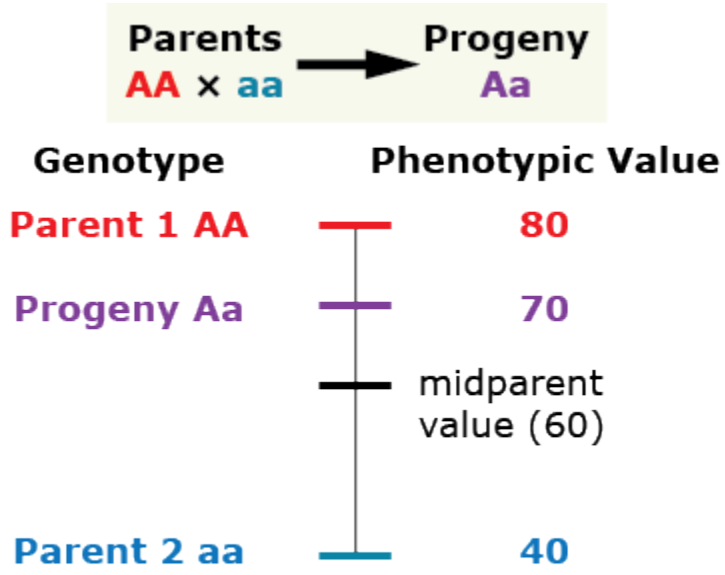


Fig. 16

Over-dominance

The phenotype of the heterozygous progeny is greater than either parent.

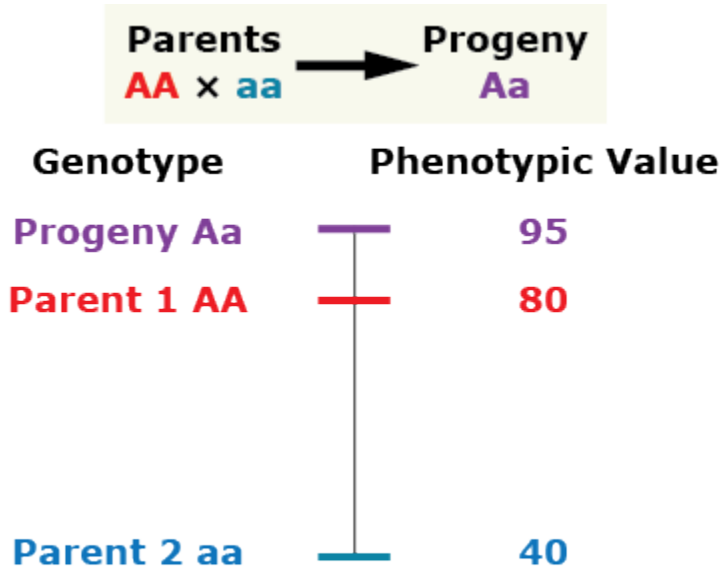


Fig. 17

Deviations from Expected Phenotypes

Multiple Alleles

With complete dominance of the type that we have been discussing, two different alleles exist for a trait, but only one of the alleles is observed in the phenotype. But it is important to understand that dominance does not affect the way in which genes are inherited. For some characters, there are reasons other than dominance among alleles at the same locus that explain deviations from expected phenotypes.

Multiple alleles—rather than just two—can occur at a single locus. Examples of multiple alleles at a single locus include the **ABO** blood group system in humans or the **S** alleles that control self-incompatibility in plants. Multiple alleles at a locus are sometimes referred to as an allelic series. However, while there may be more than two alleles per gene present in a **population**, be aware that the genotype of any given **individual** diploid plant in the population possesses only two alleles.

Penetrance is a measure of the percentage of individuals having a particular genotype that express the expected phenotype. Incomplete penetrance occurs when a genotype does not always produce the expected phenotype.

Expressivity is a related concept that describes the degree to which a character is expressed.

Incomplete Penetrance

Incomplete penetrance and variable expressivity are due to effects of other genes or environmental factors that change the effect of a particular gene. For example, a phenotype produced by an enzyme encoded by a particular gene may be expressed only within a narrow temperature range. In barley, a recessive allele occurs that produces albino plants when they are grown at lower temperatures. The allele inhibits chlorophyll production. But if barley plants that are homozygous recessive for this allele are grown above a critical temperature, the effect is not present so the plants have normal chlorophyll and are green.

Lethal alleles can change expected phenotypic ratios as well. Lethal alleles cause death when present, so that one or more genotypes will be missing from the offspring of a cross. Lethal alleles can be recessive (causing death only in homozygotes) or dominant (both homozygotes

and heterozygotes with the allele will die). Dominant lethal alleles are rarely maintained in populations.

Essential genes are genes that when mutated can result in a lethal phenotype.

Mendelian Heredity

Gregor Mendel analyzed the segregation of hereditary traits. We now know that the genotype is the genetic constitution of an organism and the phenotype is the observable characteristic or set of characteristics of an organism produced by interactions between its genotype and the environment. The phenotype is influenced by not only the genotype but also environmental effects and developmental events and by actions of other genes and their products. Therefore, individuals with the same genotype can have different phenotypes and conversely, individuals with the same phenotype can have different genotypes.

Terminology

The parental generation of a cross is often called the **P generation**. Using symbolism based on what is called the **F Symbol**, the progeny of the mating of two parents is typically called the **F_1 or first filial generation.** The subsequent generation produced by either self-pollination or crossing among the F_1 offspring (a type of mating called **inbreeding**) is referred to as the **F_2 generation**, or the **second filial generation**. The progeny resulting from self-pollination of each consecutive generation following the F_2 is referred to as F_3, F_4, F_5, and so on. Another kind of symbolism is based on the **S Symbol**. The S symbol is used to describe the offspring of a single cross—specifically the cross between two homozygous parents. F and S symbolism have been developed to describe progeny developed by hybridization and self-pollination.

F and S Symbolism

It is important to note on pages 28-33 of Fehr's textbook, plant breeders have developed a variety of systems using either the F or the S symbol to describe progeny developed by hybridization and self-pollination. What is challenging is that depending on the plant breeder, F and S symbols may be used in different, often contradictory, ways. The table below depicts examples of the particular system chosen and the way in which symbols are defined for use (Fehr, 1987, p. 28-33).

Table 3

Symbol	Description
F_1	Hybrids produced from the mating of homozygous parents.
$F_2 = S_0$	First segregating generation produced from the cross of two or more parents
$F_3 = S_1$	Offspring from self-pollination of F_2 (or S_0) plants
$F_5 = S_3$	Offspring from self-pollination of F_4 (or S_2) plants
Syn $_1$	Synthetic $_1$ = Offspring from random mating of an F_2 population
Syn $_4$	Synthetic $_4$ = Offspring from random mating of a Syn_3 population
$F_{2:5}$ line	F_2-derived line in F_5 = an F_5 generation line available for planting that originated from an F_2 generation
$S_{2:9}$ line	S_2 -derived line in S_9 = an S_9 generation line available fro planting that originated from an S_2 generation

Crosses

A cross involving a single trait (e.g., seed color) is referred to as a **monohybrid** cross, while one involving two traits (e.g., seed color and plant height) is termed a **dihybrid** cross. Conventionally, in equations used to symbolize a cross, the female parent is listed first and the male parent second, as in this example involving a single locus in diploid individuals:

AA x aa ⇒ Aa

Crosses that are done both ways are referred to as **reciprocal crosses**. For example, the reciprocal cross of the one above would be:

aa x AA ⇒ Aa

Reciprocal crosses can be used to determine whether a trait is maternally inherited. If a trait is controlled by genes located in cytoplasmic DNA, the segregation ratios between reciprocal crosses would be different because cytoplasmic DNA is inherited only through the female parent.

Predicting Segregation Ratios

If the genetic basis of a trait is known, principles developed by Mendel can be used to predict the outcome of crosses. There are three common approaches used to analyze segregation results, two of which use the listing of all possible genotypes and phenotypes of zygotes and gametes by systematic enumeration and the other of which uses mathematical rules.

- The Punnett Square Method is best for situations involving one or two genes. All possible gametes are written down in a square and then combined systematically to depict an array of genotypes of the offspring.
- The Branching or Forked-Line Method [See Appendix C for some examples] also works well for situations involving one or two genes. It uses a tally system in a diagram of branching lines.
- The **Probability Method** is based on two rules in mathematical probability theory—the Multiplicative Rule and the Additive Rule—and deals with the frequency of events.

Punnett Square Examples

Parental Monohybrid Cross

Trait	Seed color
Alleles	Y yellow y green
Cross	yellow seeds x green seeds YY x yy (homozygous dominant x homozygous recessive)
Offspring called	F_1 generation
Genotype	all alike Yy (heterozygous)
Phenotype	all alike Yy (green)

Results:

		Pollen	
Egg		**1/2y**	**1/2y**
	1/2Y	1/4Yy	1/4Yy
	1/2Y	1/4Yy	1/4Yy

F_1 Monohybrid Cross

Alleles	Y yellow y green
Cross	yellow seeds x green seeds Yy x Yy (heterozygous x heterozygous)
Offspring called	F_2 generation
Genotypic ratio	1:2:1 YY (homozygous dominant): Yy (heterozygous): yy (homozygous recessive)
Phenotypic ratio	3:1 Y_ (yellow): yy (green)

Results:

	Pollen		
Egg		**1/2Y**	**1/2y**
	1/2Y	1/4 YY	1/4 Yy
	1/2y	1/4 Yy	1/4 yy

Dihybrid Cross

Trait	Seed shape and seed color
Alleles	R round, r wrinkled, Y yellow, y green
Cross	Round, yellow seeds x round, yellow seeds RrYy x RrYy (heterozygous x heterozygous)
Offspring called	F_3 generation
Genotypic ratio	1:2:1:2:4:2:1:2:1 RRYY:RRYy:RRyy:RrYY:RrYy:Rryy:rrYY:rrYy:rryy
Phenotypic ratio	9:3:3:1 R_Y_ (round, yellow): R_yy (round, green): rrY_ (wrinkled, yellow): rryy (wrinkled, green)

Results:

	Pollen				
Egg		**1/4 RY**	**1/4 Ry**	**1/4 rY**	**1/4 ry**
	1/4 RY	1/16 RRYY	1/16 RRYy	1/16 RrYY	1/16 RrYy
	1/4 Ry	1/16 RRYy	1/16 RRyy	1/16 RrYy	1/16 Rryy
	1/4 rY	1/16 RrYY	1/16 RrYy	1/16 rrYy	1/16 rrYy
	1/4 ry	1/16 RrYy	1/16 Rryy	1/16 rrYy	1/16 rryy

Branching or Forked-Line Method

Below is an example of the forked-line or branch diagram method for determining the outcome of an intercross involving three independently assorting genes in peas.

Traits	Plant height, seed color and seed texture
Alleles	D tall / d dwarf G yellow / g green W round / w wrinkled
Cross	Tall plants with yellow, round seeds x dwarf plants with green, wrinkled seeds DDGGWW x ddggww (homozygous dominant x homozygous recessive)
F_1	DdGgWw

Expected F_2 phenotypes for each trait

Segregation of gene for plant height	**Segregation of gene for seed color**	**Segregation of gene for seed texture**	**Combined phenotype of all three genes**	
3/4 D_(tall)	3/4 G_(yellow)	3/4 W_(round)	⇒	27/64 D_G_W(tall, yellow, round)
		1/4 ww(winkled)	⇒	9/64 D_G_ww(tall, yellow, wrinkled)
	1/4 gg(green)	3/4 W_(round)	⇒	9/64 D_ggW_(tall, green, round)
		1/4 ww(wrinkled)	⇒	3/64 D_ggww(tall, green, wrinkled)
1/4 dd(dwarf)	3/4 G_(yellow)	3/4 W_(round)	⇒	9/64 ddG_W_(dwarf, yellow, round)
		1/4 ww(wrinkled)	⇒	3/64 ddG_ww(dwarf, yellow, wrinkled)
	1/4gg(green)	3/4 W_(round)	⇒	3/64 ddggW_(dwarf, green, round)
		1/4 ww(wrinkled)	⇒	1/64 ddggww(dwarf, green, wrinkled)

Rules of Probability

Using probability theory can allow for accounting of the frequency of events, such as the chance of obtaining a head on a coin toss or obtaining a dominant homozygote (**AA**) from the mating between two heterozygotes (**Aa**). To figure out the probability of an event, all possible outcomes must be determined. For a coin toss, there are two possible events—heads or tails—each with a probability of ½ that it would occur. For the progeny produced by a heterozygote, the probability associated with each type of offspring is ¼ (**AA**), ½ (**Aa**) and ¼ (**aa**).

The Multiplicative Rule states that if events X and Y are independent, the probability that they occur together (that is A and B), is the probability of A times the probability of B. It is denoted as:

$$P(A) \times P(B)$$

The Additive Rule states that if events X and Y are independent, the probability that at least one of them occurs (that is A or B), is the probability of A plus the probability of B minus the probability that both A and B occur together. It is denoted as:

$$P(A) + P(B) - [P(A) \times P(B)]$$

Mendel's Principles

Mendel's analysis of monohybrid crosses identified three key principles:

The Principle of Uniformity

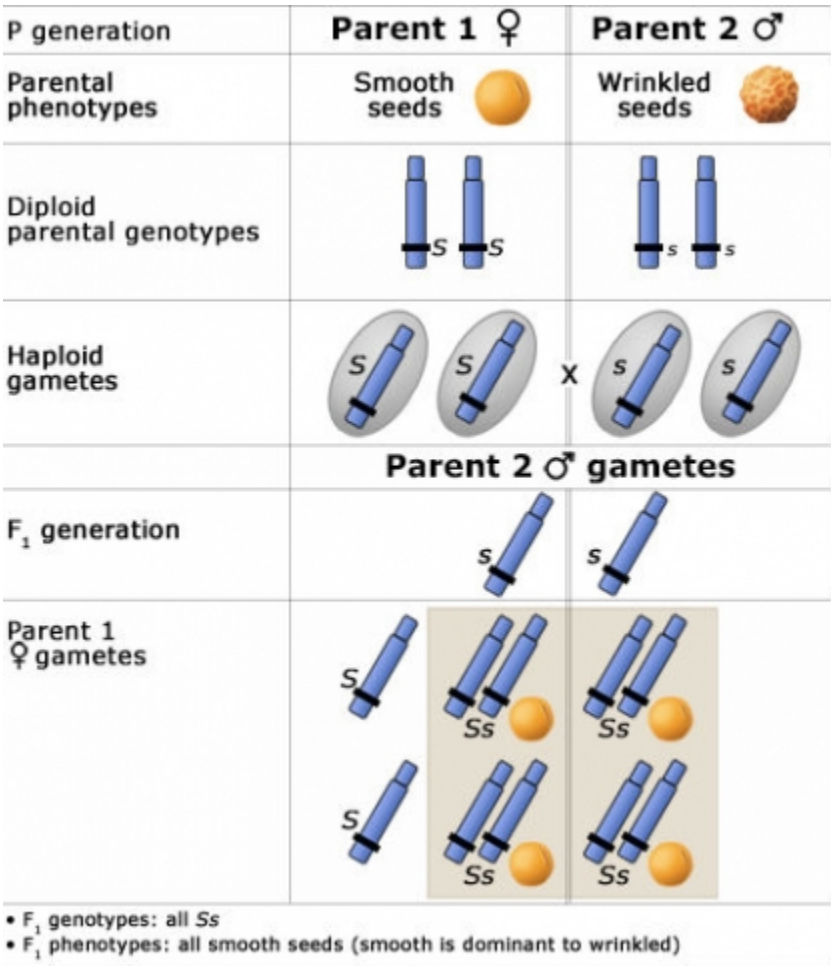


Fig. 18 Punnett Square showing the Principle of Uniformity.

If both parents are homozygous, their F_1 is genetically uniform.

To the right is a Punnett Square showing an example of this phenomenon, depicting the genotypic and phenotypic ratios and chromosomes of the diploid parents, haploid gametes, and the F_1 generation.

The Principle of Segregation

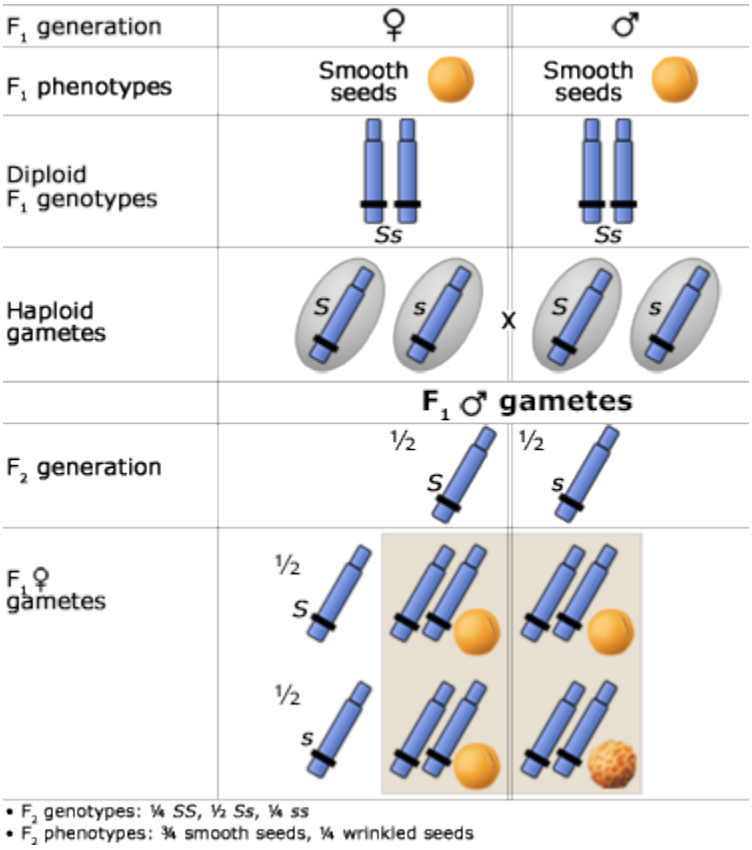


Fig. 19 Punnett Square showing the Principle of Segregation.

In a heterozygote, two different alleles of a gene locus segregate from each other in the formation of gametes. Below are two figures (one using a Punnett Square and the other the fork or branch diagram method) showing an example of Mendel's law of segregation. The figures depict the genotypic and phenotypic ratios and chromosomes of the F_1 heterozygote, haploid gametes, and the F_2 generation.

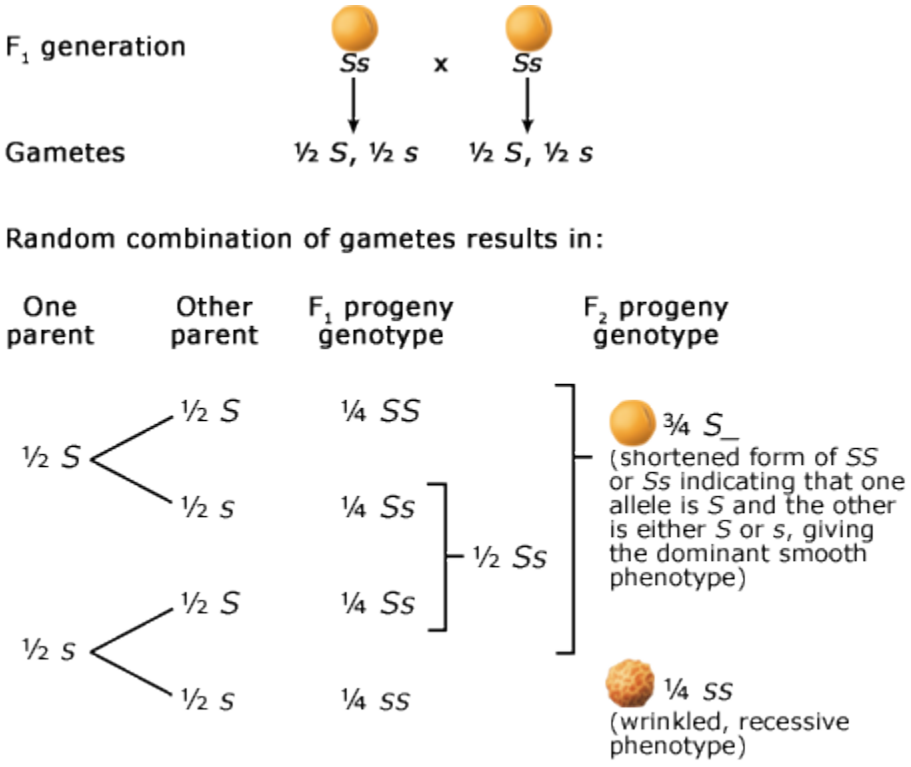


Fig. 20 Fork or Branch Diagram showing the Principle of Segregation.

The Principle of Independent Assortment

Alleles at different gene loci are transmitted independently of one another during the production of gametes. Below are two figures (one using a Punnett Square and the other the fork or branch diagram method) showing an example of Mendel's law of independent assortment. The figures depict the genotypic and phenotypic ratios and chromosomes of the parents, the F_1 heterozygote, haploid gametes, and the F_2 generation.

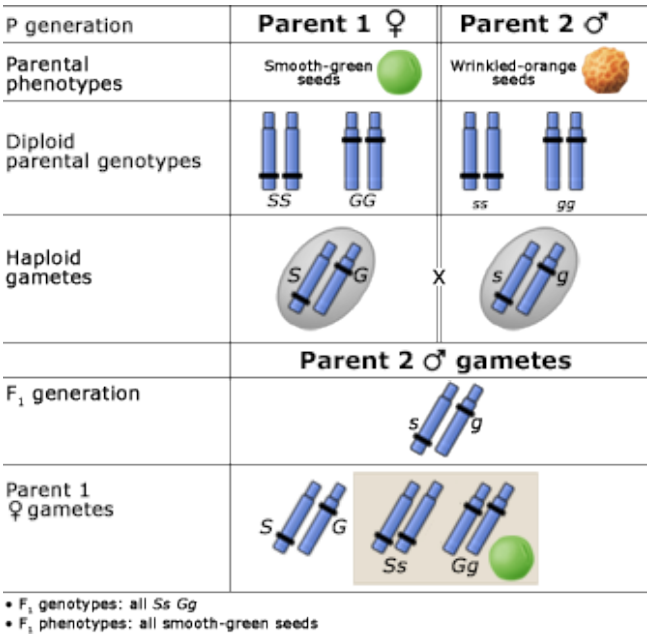


Fig. 21 Punnett Square showing the Principle of Independent Assortment.

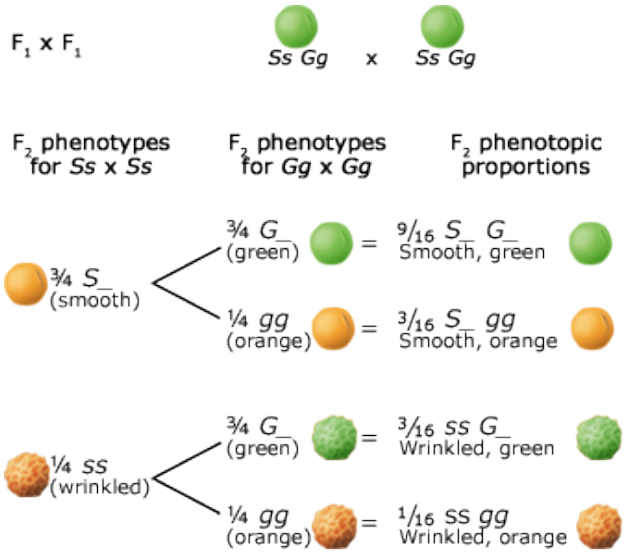


Fig. 22 Fork or Branch Diagram showing the Principle of Independent Assortment.

Inheritance

A trait or characteristic may be under the control of one or more genes. The range of variation for a particular characteristic indicates the mode of inheritance of that characteristic.

- **Qualitative inheritance** — simple inheritance of a characteristic under the control of single gene or a few major genes. The expression of simply inherited characteristics is discrete. That is, the phenotypic variation of the characteristic can be separated into distinct classes. Generally, the environment has little influence on the characteristic's expression.
- **Quantitative inheritance** — inheritance of characteristics influenced by numerous genes (multiple genes or polygenes). The involved genes have small, cumulative effects on the phenotype of the characteristic. The expression of such characteristics can be measured in quantitative units that are continuous, rather than discrete, and is often considerably influenced by the environment. Quantitative inheritance is the subject of the module on Inheritance of Quantitative Traits.

The inheritance of some characteristics cannot easily be categorized as either qualitative or quantitative. These characteristics are usually under the control of one or few major genes

as modified by multiple genes with small effects. Together with environmental effects, the phenotype of such characteristics may show continuous variation.

Progeny Ratios

To determine the mode of inheritance of a particular character, plant breeders mate plants and evaluate the performance of their offspring. The proportion of progeny exhibiting different phenotypes provides information about the proportion of progeny possessing different genotypes.

- **Phenotypic ratio** — the proportion of progeny exhibiting different phenotypes
- **Genotypic ratio** — the proportion of progeny possessing different genotypes

These ratios are commonly determined by crossing two plants having contrasting phenotypes for a given character. The parents may or may not be homozygous. The progeny are heterozygous for the trait. Self-pollinating the F_1 progeny produces the F_2 generation, and so forth (F_n). In each generation, the ratio of plants displaying contrasting phenotypes for the particular trait reveals information about the genotypes of the parents, as well as gene action (e.g., dominant or recessive alleles).

In the exercise concerning phenotypic and genotypic ratios, with each consecutive generation, the proportion of heterozygotes (**Gg**) is reduced. With continued self-pollination, the heterozygotes will segregate, decreasing the proportion of heterozygotes in the population by half each generation. Notice that the homozygotes can only produce homozygotes.

Successive Generations

Table 1

Generation	Heterozygosity (%)
F_1	100.0
F_2	50.
F_3	25.0
F_4	12.5
F_5	6.25
F_6	3.12

For each successive generation of offspring resulting from one F_1 individual, by the F_8 generation, the population is essentially homozygous. When no further segregation for the trait occurs, all progeny derived from that F_1 will "breed true" because they are homozygous for the trait.

The proportion of plants that are expected to be heterozygous at any gene when starting with a heterozygous F_1 and selfing can be determined by using the formula $(½)^n$, where n = the number of segregating generations, e.g., in F_2, n = 1 and in F5, n = 4. Using this we get the following proportions of heterozygous plants in F_4: $(½)^n = (½)^3 = ⅛ = 12.5\%$.

The proportion of homozygous plants in any generation is then given by $1 - (½)^n$ which, when algebraically converted, is equal to:

$$\frac{2^n - 1}{2^n}.$$

Applying this to F_4 we get $\frac{2^3 - 1}{2^3} = \frac{8 - 1}{8} = \frac{7}{8} = 87.5\%$

When working with actual genotypes we must remember that in any segregating generation there are two homozygous genotypes and we expect equal quantities of each. Using the example of an F_1 that is **Aa**, in F_2, we expect ¼ **AA** + ½ **Aa** + ¼ **aa**.

In F_4 we expect to be homozygous with half of those **AA** and half **aa**. Thus overall we expect the following F_4 genotypic frequencies:

$$\frac{7}{16}\mathrm{AA} + \frac{1}{8}\mathrm{aa} + \frac{7}{16}\mathrm{aa}$$

Scenarios under cross-pollination — with and without selection — will be discussed in more detail in the module on Population Genetics.

Progeny Test

There are two principal procedures that allow the plant breeder to determine the basis of phenotypes (genetic or environmental), gene action, and the genotypes of individual plants. Which procedure is used depends on the specific objectives of the breeder.

Progeny Test

The **progeny test** evaluates the genotype of an individual based on the performance of its offspring. The progeny test can be used to:

1. Distinguish heritable phenotypes from phenotypes attributable to environmental effects.
2. Determine the genotype or the allelic composition of an individual.

Steps in Progeny Test

1. Hybridize (mate) two plants, A and B.

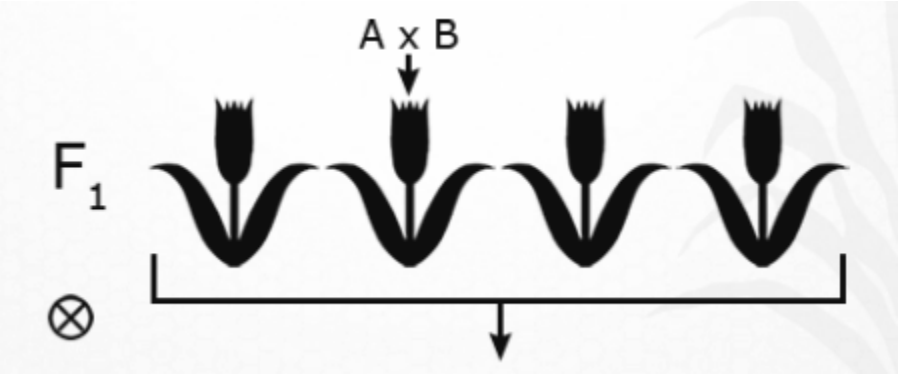


Fig.23 Hyridization

2. Grow out and self-pollinate the F_1 plants.

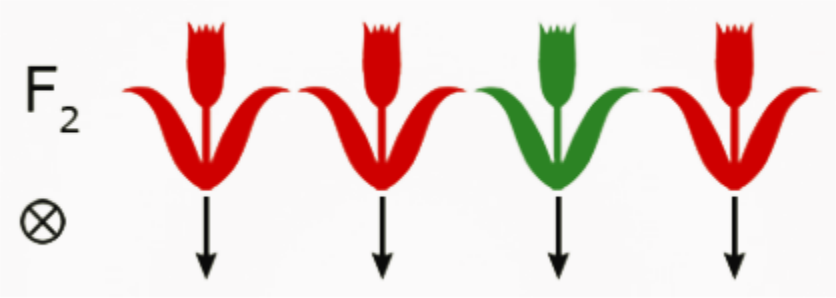


Fig.24 Self-pollination.

3. Grow out and self-pollinate F_2 plants.
    a. Determine the phenotypic ratio of trait(s) of interest.
    b. Harvest seed separately from each plant.
4. Plant a portion of the F_3 seed from each phenotype separately.
    a. Determine the phenotypic ratio in each group—the phenotypic ratio reveals which of the F_2 plants were homozygous and which were heterozygous for the trait(s) of interest.
    b. Based on the phenotype information, calculate the genotypic ratio.

In this example, the phenotypic ratios of the F_3 plants reveal the following genotypic information about each of the F_2 parents:

F_2 Parent	**Genotype**
a	Homozygous red
b	Heterozygous red
c	Homozygous green
d	Heterozygous red

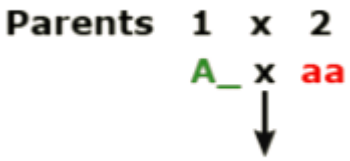
Both the red and green phenotypes occur in ratios consistent with those of heritable traits. Thus, there is a genetic basis for these phenotypes (i.e., these phenotypes are not just the result of environmental conditions).

Testcross

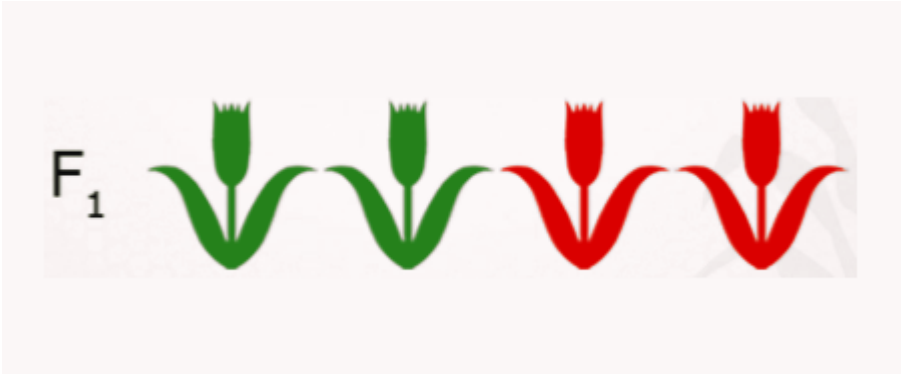
The **testcross** procedure is used to determine the genotype of an individual or **linkage** groups. Linkage is a condition in which genes located on the same chromosome are inherited together due to their close proximity. Linkage will be discussed in greater detail in "Linkage" module.

Steps in Testcross

1. Hybridize (mate) two plants. The genotype of Parent 1 is unknown, A (?). Parent 2 is homozygous recessive for the trait of interest, aa.



2. Grow out F_1 plants and evaluate the phenotypic ratio:
   a. If segregating 1:1, then you know that the genotype of Parent 1 was heterozygous, Aa.



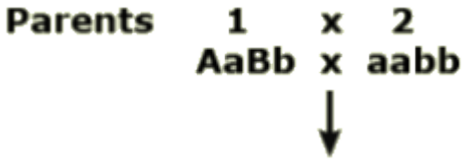
   b. If all plants have the phenotype of Parent 1, than you know that Parent 1 was homozygous dominant, AA.

The **backcross** is a special type of progeny test. It is a cross of an F_1 to either of the original

parents. This procedure is used extensively in basic genetic studies but not often used by plant breeders to determine genotypes of plants.

Determine Linkage

To determine linkage groups, hybridize two plants:



- Parent 1 is heterozygous at two (or more) loci.
- Parent 2 is homozygous recessive at these loci.

The interpretation of the results of this cross will be discussed in the module on Linkage.

Genetic Recombination and Its Effects

Develop Improved Cultivars

To develop improved cultivars, plant breeders usually combine the favorable characteristics of one plant or cultivar with the desirable traits of another plant or cultivar, accumulating desirable alleles for key characters. To obtain an improved genetic combination, breeders make a series of matings, selecting the best offspring to produce the next generation. Plant breeders rely on several genetic mechanisms to obtain new genetic combinations.

1. **Segregation** — Homologous chromosomes derived from different parents separate and distribute randomly to cells during meiosis.
2. **Recombination** — Formation of new gene combinations by mating individuals having differing genotypes.

Segregation

Segregation is the result of the independent assortment or chance distribution of homologous chromosomes and the genes that they carry to gametes. Through meiosis, allelic pairs are separated and distributed to different cells, which subsequently undergo gametogenesis.

Genes located on different chromosome pairs assort independently. That is, the chance

distribution of a particular chromosome, say one of these green chromosomes, to one cell, has no effect on the distribution of a yellow chromosome. Independent assortment facilitates recombination and leads to segregation in subsequent generations.

Recombination

Mating two plants possessing different genotypes results in progeny with genotypes that may differ from the parental types. The progeny having genotypes that differ from the parents are referred to as "**recombinants**."

Restrictions with Independent Assortment

Hybrid Characteristics

A breeder cannot improve a characteristic unless there is some variability for that characteristic within which to make selections. Hybridizing plants differing in their phenotypes (and genotypes) and selecting from among the recombinants provide the breeder with the opportunity to make progress towards crop improvement. However, recombination and segregation may fail to provide the expected variation for two general reasons.

- **Population size** — A minimum of progeny from a cross must be grown out and evaluated. If the number is too small, the likelihood of the desired recombinant occurring in the population is reduced. As the number of independently assorting genes increases, the number of plants that must be evaluated increases exponentially. Thus, an adequate population is essential to make efficient progress towards the breeding goals. The minimum population size required for all genotypes to be represented in the population can be calculated as follows:

    1. Determine the number of segregating gene pairs. Let that number equal "n".
    2. Calculate the minimum population size: minimum population size = 4^n

- **Gene Interaction** — Although the genes involved in **epistatic** and **pleiotropic** interactions may assort independently, their interactions often affect phenotypic and genotypic ratios.
- **Linkage** — As stated earlier, loci in close proximity on the same chromosome tend to be

transmitted together and do not assort independently.

Genetic Cross-Data

When analyzing data from genetic crosses, it is frequently appropriate to use some kind of statistical analysis because such data is often quantitative. One statistical procedure commonly used for testing results of segregation data is called a **chi-square (χ^2) test.** The chi-square test is also known as a "goodness-of-fit" test.

Breeders wonder if data support or fit a particular hypothesis and therefore help to explain the results. For example, does the range of phenotypes observed within the progeny of a cross-fit a particular segregation ratio, e.g., 3:1 or 9:3:3:1? The chi-square procedure helps breeders understand the significance of **deviation of observed results from results predicted** by the hypothesis being tested. A **null hypothesis** is formed that states there is no real difference between the observed and expected data. If differences are due to chance, then the hypothesis can be accepted, otherwise, the null hypothesis is rejected and the breeder can modify the hypothesis in favor of a better one. The equation used to calculate the (χ^2) statistics is as follows

$$x^2 = \sum \frac{(\text{observed} - \text{expected})^2}{expected}$$

The chi-square procedure will be covered in more detail in the Quantitative Methods course.

Gene Interactions

Traits

When multiple genes control a particular trait or set of traits, gene interactions can occur. Generally, such interactions are detected when genetic ratios deviate from common phenotypic or genotypic proportions.

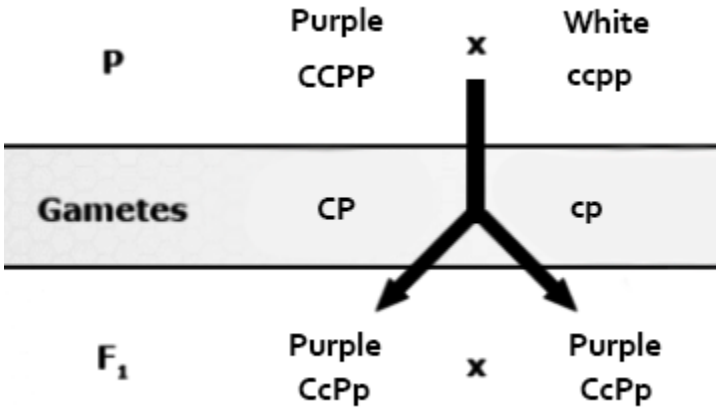
- **Pleiotropy** — Genes that affect the expression of more than one character
- Epistasis**epistasis** — Genes at different loci interact, affecting the same phenotypic trait. Epistasis occurs whenever two or more loci **interact to create new phenotypes**. Epistasis also occurs whenever an allele at one locus either **masks** the effects of alleles at one or more loci or if an allele at one locus **modifies** the effects of alleles at one or more loci.

There are numerous types of epistatic interactions.

Epistasis is expressed at the **phenotypic** level. It is important to note that genes that are involved in an epistatic interaction may still exhibit independent assortment at the **genotypic** level. The following slides show some examples of epistasis drawn from various types of plants.

Duplicate Recessive Epistasis

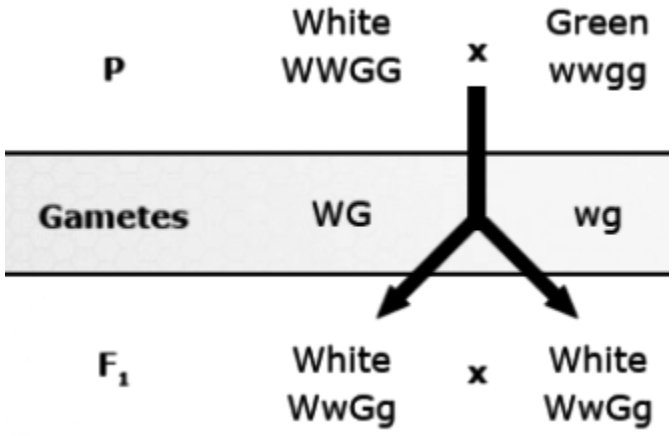
Duplicate recessive epistasis (also known as **complementary action**): 9:7 ratio observed in flower color of progeny of crosses between a pure line pea plant with purple flowers (genotype **CCPP**) with a pure line, homozygous recessive plant with white flowers (**ccpp**). The F_1 plants are all purple and have a genotype of **CcPp**, but the F_2 progeny will have a modified ratio of 9:7 because color is only produced if both genes have at least one dominant allele. These genes control flower color by controlling the expression of biochemical compounds known as anthocyanins that impart pigment to the flower. Pigmentation in this case is controlled by a two-step chemical reaction. One of these genes controls the first step and the other controls the second step.



		Male Gametes			
		CP	Cp	cP	cp
Female Gametes	CP	CCPP Purple	CCPp Purple	CcPP Purple	CcPp Purple
	Cp	CCPp Purple	CCpp White	CcPp Purple	Ccpp White
	cP	CcPP Purple	CcPp Purple	ccPP White	ccPp White
	cp	CcPp Purple	Ccpp White	ccPp White	ccpp White

Dominant Epistasis

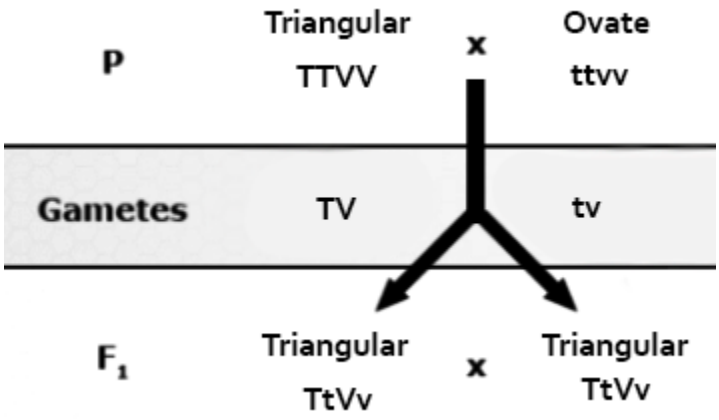
Dominant epistasis (also known as **masking action**): 12:3:1 ratio observed in fruit color of progeny of crosses of squash. In the F_2, fruits are white if the genotypes are either **W_G_**or **W_gg** because the dominant allele for the first gene (**W**) masks the effect of either allele for the other gene (**G** or **g**). Color is present only if the first gene is homozygous recessive (**ww**). Yellow squash have the genotype **wwG_** and green ones have the genotype **wwgg**.



		Male Gametes			
Female Gametes		WG	Wg	wG	wg
	WG	WWGG White	WWGg White	WwGG White	WwGg White
	Wg	WWGg White	WWgg White	WwGg White	WwGg Yellow
	wG	WwGG White	WwGg White	wwGG Yellow	wwGg Yellow
	wg	WwGg White	Wwgg White	wwGg Yellow	wwgg Green

Duplicate Dominant Epistasis

Duplicate dominant epistasis (also known as **duplicate action**): 15:1 ratio observed in fruit shape of progeny of crosses of the common shepherds purse. If either of the two genes involved in fruit shape (**T** or **V**) are present alone or both together (**TV**), then the plants will all produce triangular-shaped fruit. Only the homozygous recessive genotype (**ttvv**) produces a seed capsule with an ovate shape.



	Male Gametes				
Female Gametes		**TV**	**Tv**	**tV**	**tv**
	TV	TTVV Triangular	TTVv Triangular	TtVV Triangular	TtVv Triangular
	Tv	TTVv Triangular	TTvv Triangular	TtVv Triangular	Ttvv Triangular
	tV	TtVV Triangular	TtVv Triangular	ttVV Triangular	ttVv Triangular
	tv	TtVv Triangular	Ttvv Triangular	ttVv Triangular	ttvv Ovate

References

Hartwell, L., L. Hood, M. Goldberg, A. Reynolds, and L. Silver. 2011. Genetics: From Genes to Genomes. 4th edition. McGraw-Hill Publ., Boston, MA.

National Institutes of Health. National Human Genome Research Institute. "Talking Glossary of Genetic Terms." http://www.genome.gov/glossary/

Pierce, B. 2010. Genetics: A Conceptual Approach. 4th edition. W. H. Freeman, NY.

Pierce, B. A. 2008. Genetics: A Conceptual Approach. 3rd edition. W. H. Freeman, NY.

Russell, P. J. 2010. iGenetics: A Molecular Approach. 3rd edition. Benjamin Cummings, San Francisco, CA

How to cite this chapter: Muenchrath, D., A. Campbell, L. Merrick, T. Lübberstedt, and S. Fei. (2023). Gene segregation and genetic recombination. In W. P. Suza, & K. R. Lamkey (Eds.), *Crop Genetics*. Iowa State University Digital Press. DOI: 10.31274/isudp.2023.130

Linkage

Thomas Lübberstedt; Arden Campbell; Deborah Muenchrath; Laura Merrick; and Shui-Zhang Fei

Introduction

Genes located on the same chromosome are genetically linked. Genetic linkage analysis can be used to determine the order of genes on chromosomes. Closely linked genes are not segregating independently, like genes located on different chromosomes. This has different implications, e.g., in relation to trait correlations. Moreover, linked genes can be used as genetic markers, which have become an important tool in plant breeding.

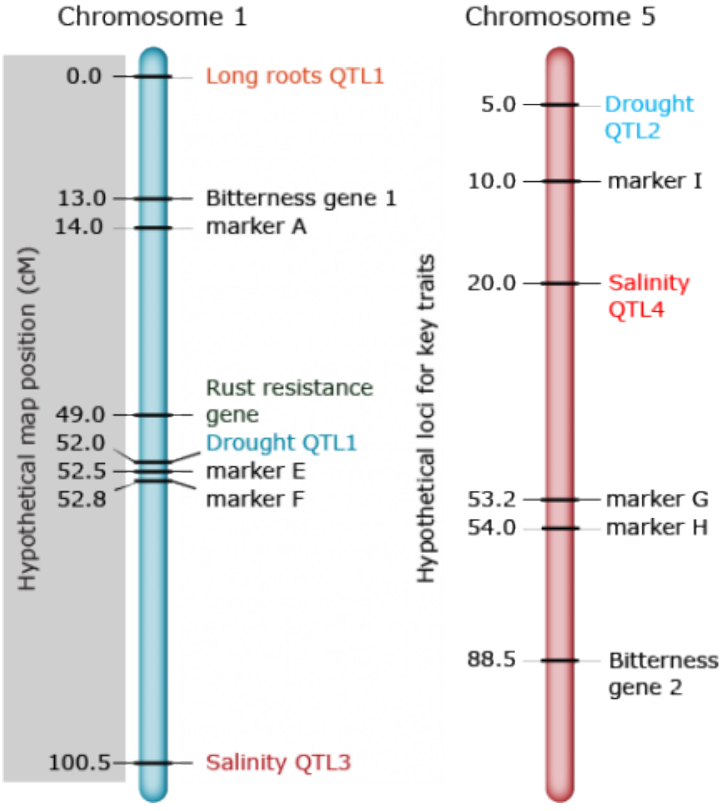


Fig. 1 Genes located on the same chromosome are genetically linked. Illustration by Walter Suza.

Objectives

- Develop an understanding of the genetic basis of linkage.
- Gain awareness on how to detect the occurrence of linkage.
- Review the principles of genetic map construction.
- Become familiar with the concept of linkage disequilibrium.

Crossover and Recombination

Genetic Organization

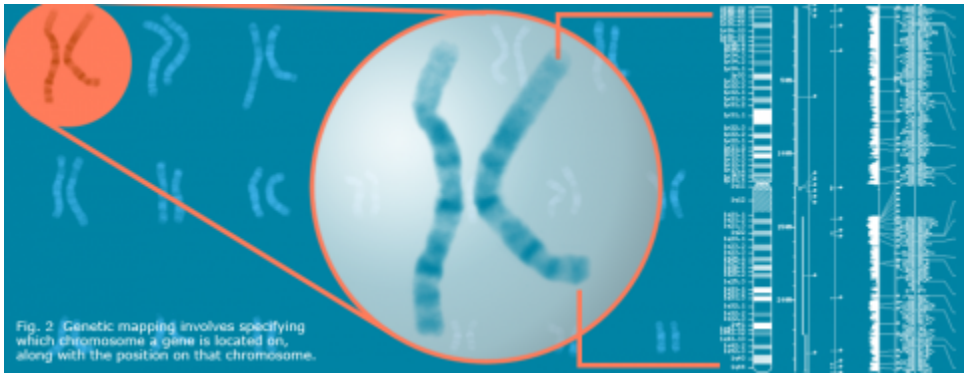


Fig. 2 Genetic mapping involves specifying which chromosome a gene is located on, along with the position on that chromosome. Illustration by Iowa State University.

Genes are physically organized on chromosomes. Each gene is located at a particular "address" (particular position on a specific chromosome, which can be identified by genetic mapping). Inheritance of genes located on different chromosomes follows the rules of independent assortment. Since plant species have multiple chromosomes, independent assortment is true for the majority of genes. In contrast, linked genes located on the same chromosome are more likely to cosegregate, i.e., being jointly transmitted to offspring more often than expected by independent assortment. The biological process that separates linked genes is the crossing-over (C.O., or crossover), which occurs during meiosis, and leads to genetic recombination.

Crossing-Over

During meiosis of diploid organisms, the chromatids of homologous chromosomes pair and form bivalents. During Meiosis I, homologous chromatids pair to physically exchange chromosome segments. The chromosomal site, where this reciprocal exchange of homologous chromosome segments takes place, is called a **chiasma**. Thus, crossing-over involves not completely understood mechanisms for identification of homologous sites of chromatids, breakage and rejoining of chromosomes.

Genetic Distance

Crossing-over events occur more or less random during meiosis. In most plant species, one to few crossing-over events occur per meiosis and chromosome. Thus, the closer the genes are physically linked on the same chromosome, the less likely they will get separated, and consequently, the less likely genetically recombinant gametes will be produced. This is the underlying principle of genetic maps: the genetic distance between genes reflects the probability of a crossing-over between linked genes.

Recombination

Observation of crossing-over events requires cytological methods, which can be cumbersome for large populations. In contrast, genetic recombinants can be observed at the phenotype level, or by use of DNA markers. If two linked genes with two alleles each have clear phenotypic effects, e.g., on flower color (**A**: red, **a**: white; **A** is dominant over **a**) and seed color (**B**: green, **b**: yellow; **B** is dominant over **b**), then genetic recombinants can easily be identified by determining the fraction of non-parental gametes in the offspring.

Note that crossing-over also takes place in meiosis of completely homozygous individuals. However, in this case, genetic recombination cannot be observed as described above. The reason is that observation of recombinant gametes requires two (or more) different alleles at the loci, for which linkage is going to be determined. This explains why offspring saved from pure line cultivars will not segregate whereas seed harvested from F_1 hybrid will segregate. The observable fraction of recombination events is also called **effective recombination**.

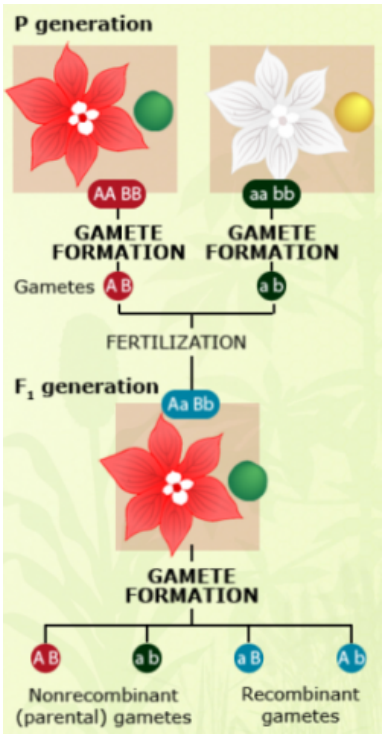


Fig. 3 Formation of gametes with parental (left side) and non-parental (recombinant; right side) allele combinations at the two loci A and B.

Linkage Detection

Linkage Phase

For linkage detection, it is crucial to know the linkage phase of alleles.

The linkage phase is the physical arrangement of linked genes in a chromosome. A double heterozygote with a genotype of **AaBb** could be in one of the two linkage phases. Conventionally, when linked dominant alleles are located on the same homologous chromosome and the linked recessive alleles are on the other homologous chromosome, for example, **AB**/**ab**, it is said the genes are linked in the coupling phase. When a dominant allele at one locus is on the same homologous chromosome as a recessive allele of the other linked gene, for example, **Ab**/**aB**, it is said that the genes are linked in repulsion phase (Fig. 4).

This knowledge is crucial, as linkage detection and distance estimation is based on the observed parental and non-parental gametes.

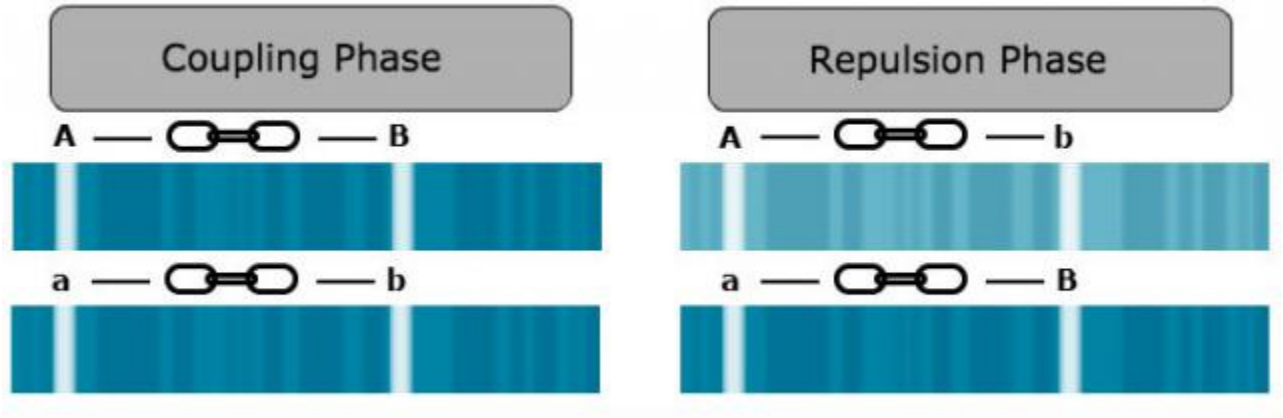


Fig. 4 Linkage phases

Coupling and Repulsion

In the case of close linkage, non-parental gametes and respective offspring are underrepresented.

An example is the Australian sheep blowfly, *Lucilia cuprina*. Normal blowflies have a green thorax and surround themselves in a brown cocoon during their pupal stage. However, recessive genes (here marked **a** and **b**) can cause the fly to develop a purple thorax and spin a black puparium.

Fig. 5 Australian sheep blowfly, *Lucilia cuprina*. Photo by fir0002, licensed under CC BY-NC via Wikimedia Commons.

Using Testcrosses

For detection of linkage, appropriate testcrosses need to be conducted. The linkage phase is known, if two homozygous parental genotypes (**AABB** and **aabb**) are crossed to produce the respective F_1 (**AaBb**).

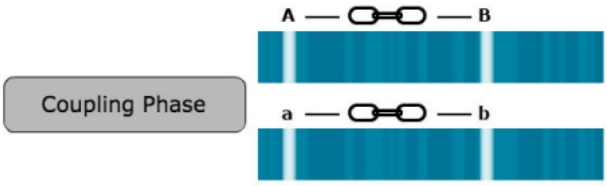


Fig. 7 Coupling phase

In this case, **A** and **B** as well as **a** and **b** are linked in coupling phase.

The non-parental recombinant gametes have the genotype **Ab** and **aB**, whereas the parental gametes have the genotype **AB** and **ab**.

Usually the phenotype cannot be observed in (haploid) gametes, but only in diploid plants. Thus, to determine whether two loci are linked, offspring need to be produced. This can be achieved by self pollination of the **AaBb** – F1, by production of doubled haploid offspring, or by a testcross.

In this particular example, a **backcross** (BC) of the F_1 to the **aabb** parent would be the best option.

Testcross Gametes

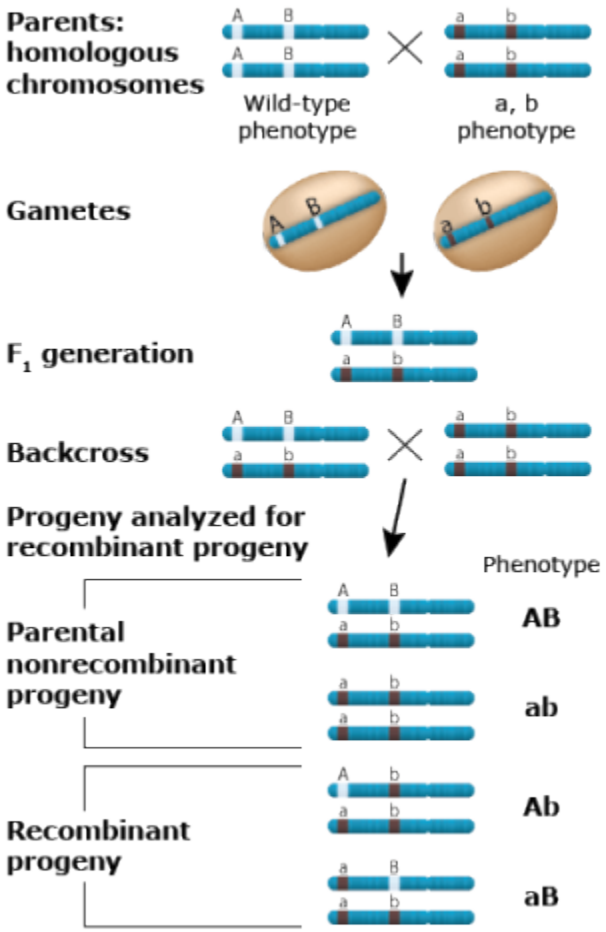


Fig. 8 Backcross depicting gene linkage

All offspring from this BC would receive an ab gamete from the **aabb** parent, and any of the two parental (**AB**, **ab**) or non-parental (**Ab**, **aB**) gametes from the F_1.

Because of dominance of **A** over **a** and **B** over **b**, all four resulting diploid genotypes in the BC_1 (backcross generation 1) generation (**AaBb**, **Aabb**, **aaBb**, **aabb**) can be phenotypically discriminated, and used to count genotypes that received parental or non-parental gametes from the F_1.

Thus, when using this BC approach, only the crossing-over events that occurred in the F_1 are monitored for linkage estimation.

Chi-Square Test

For detection of linkage, a Chi-Square test can be employed. The Chi-Square test compares observed with expected frequencies. In this case, the null hypothesis to determine expected frequencies is the assumption of independent assortment. Under this assumption, equal frequencies of all four gametes are expected. In the case of linkage, BC1 individuals carrying non-parental gametes are underrepresented, leading to a statistically significant Chi-Square value. This means that the null hypothesis of independent assortment would be rejected and linkage assumed.

Table 1 An example of the detection of linkage in Drosophila melanogaster using a Chi-Square test. d: difference between the observed number and expected number. The significantly higher Chi-Square values reject the null hypothesis and strongly indicate the presence of linkage.

Phenotypes	**Observed Number (o)**	**Expected Number (e)**	**d (o – e)**	**d^2**	**d^2 / e**
Parentals: (black-bodied and normal wing plus grey-bodied, vestigial wing)	2,712	1,618	1,094	1,196,836	739.7
Recombinants: (black-bodied and vestigial wing plus grey-bodied and normal wing)	524	1,618	1,094	1,196,836	739.7

Chi-Square Results

To better understand the use of Chi-Square in determining linkage, two numerical examples based on the cross schemes described in Figs. 8 and 9 are provided here. In both examples, a sample size of 2,000 BC1 individuals has been used.

The Chi-Square test sums up over all squared differences between observed and expected values, divided by expected values.

In example A, observed and expected values are equal, thus the Chi-Square value = 0.

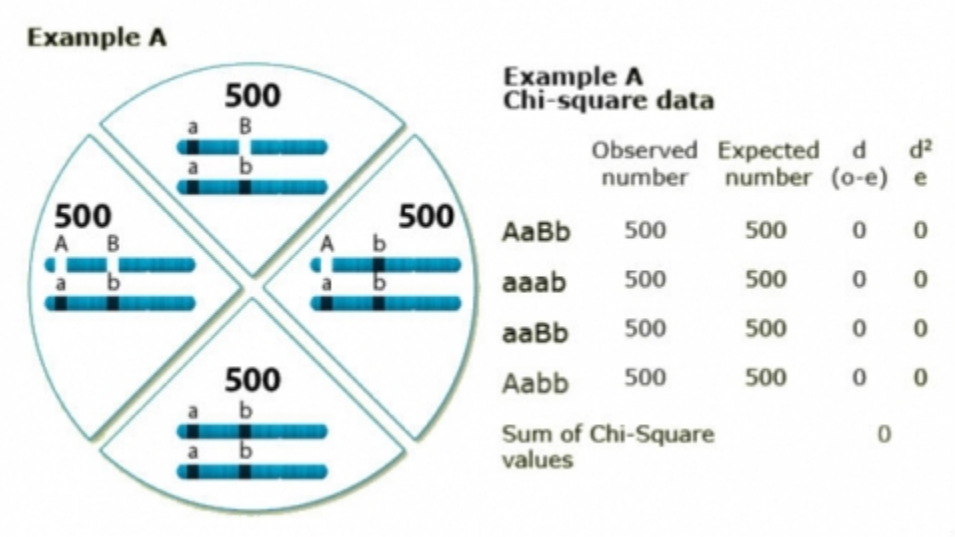


Fig. 9 Example A. Expected vs. observed distributions of phenotypes

In example B, the squared differences between observed and expected values is in all cases 90,000, to be divided by the expected 500 = 180. As there are four genotypic classes, the Chi-Square value is 720, which is significantly larger than the tabulated value of 3.81 (p = 5%).

In conclusion, example A is in agreement with independent assortment, whereas in example B, linkage has been detected.

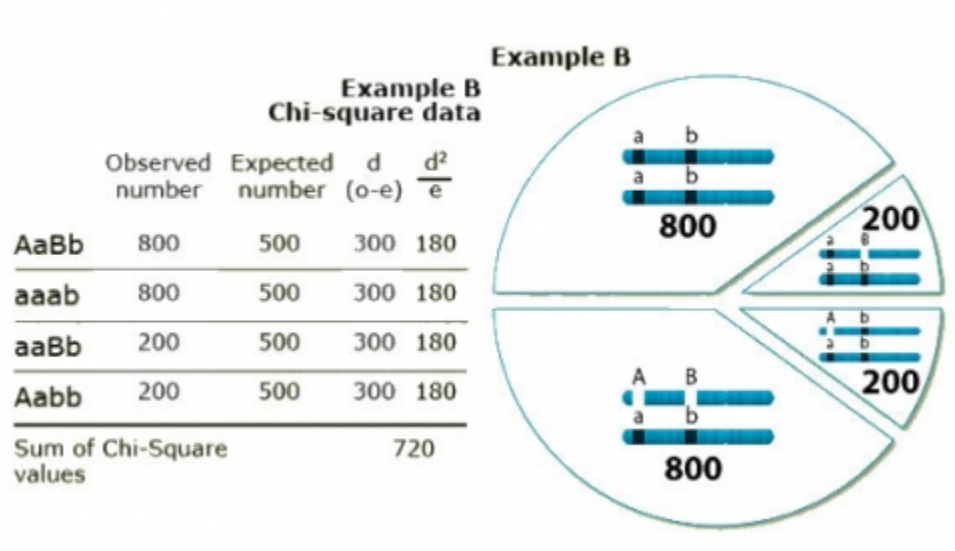


Fig. 10 Example B. Expected vs. observed distributions of phenotypes

In conclusion, example A is in agreement with independent assortment, whereas in example B, linkage has been detected.

Genetic Distance

The same data used to determine linkage can also be used to estimate the recombination frequency between two genes (more precisely, the recombinant frequency). The recombinant frequency = (number of BC1 progeny with recombinant (non-parental) alleles / total number of BC1 progeny) x 100%.

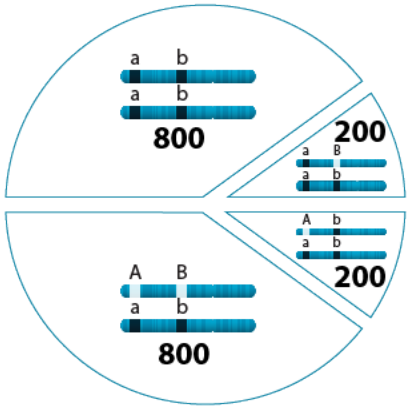


Fig. 11 200 + 200 = 400
400 / 2000 = 0.2
0.2 x 100% = 20%

In example B of Section 2: Linkage Detection, the recombinant frequency is (200 + 200 / 2000) * 100% = 20%.

The recombination frequencies between any pairs of genes provide an estimate of how close they are linked on a chromosome. The recombinant frequency in % is sometimes also called "map units" (M.U.). In this example, the genetic distance in map units between the two genes under consideration is 20 M.U.

In the case of complete linkage of two genes, no recombinants would be expected. The recombinant frequency would be 0%, which represents the lower limit of recombinant frequencies.

In the case of random segregation, the expected numbers of recombinant and nonrecombinant alleles are equal. Thus, the upper limit of recombinant frequencies in the case of unlinked or loosely linked genes is 50%.

Even for gene pairs located at the different ends of the same chromosome, recombination frequency can reach 50%. The procedure to determine recombination frequencies between any pair of genes is called two-point analysis.

Three-Point Analysis

Purpose

Whereas two-point testcrosses establish linkage between pairs of genes, three-point

testcrosses facilitate establishment of the order of genes on chromosomes, as a prerequisite to establishing genetic maps. If a third locus with alleles **C** and **c** (**C** is dominant over **c**) is added to the case mentioned in Genetic Distance, where **A** and **B** are linked in coupling phase and the dominant allele **C** is in coupling with **A** and **B**, then eight different testcross progeny would result from a backcross with the recessive parent.

Class	Genotype of gamete from heterozygous parent			Number		Origins
1	A	C	B	179	}	Parentals, no crossover
2	a	c	b	173		
3	A	c	b	52	}	Recombinants, single crossover AC
4	a	C	B	46		
5	A	C	b	22	}	Recombinants, single crossover CB
6	a	c	B	22		
7	A	c	B	4	}	Recombinants, double crossover AC, CB
8	a	C	b	2		

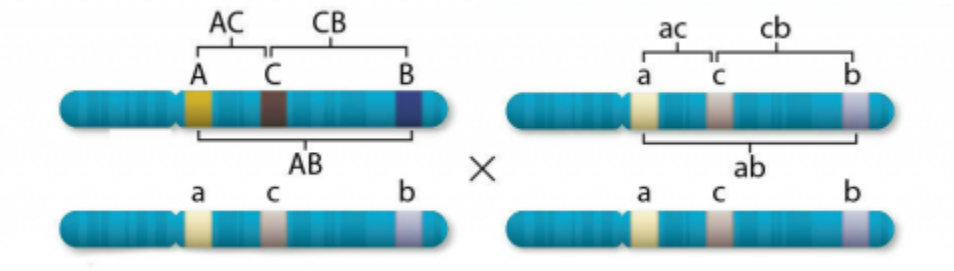


Fig. 12 Example for a testcross (backcross) as a first step toward a three-point analysis. Adapted from Russell, 2010.

Frequency Chart

Pairwise recombination frequencies can be determined as described in the chart.

- 20.8% for **AC** (**AC** recombinants are in classes 3, 4, 7, and 8; thus, the recombination rate between **A** and **C** is (52+46+4+2/500) * 100% = 20.8%)
- 10% for **CB** (**CB** recombinants are in classes 5-8)
- 28.4% for **AB** (**AB** recombinants are in classes 3-6).

Once linkage between pairs of three (or more) genes has been established, the next question is how they are arranged in linear order on chromosomes, which could be **ABC**, **ACB**, or **CAB**.

Class	Genotype of gamete from heterozygous parent			Number		Origins
1	A	C	B	179	}	Parentals, no crossover
2	a	c	b	173		
3	A	c	b	52	}	Recombinants, single crossover AC
4	a	C	B	46		
5	A	C	b	22	}	Recombinants, single crossover CB
6	a	c	B	22		
7	A	c	B	4	}	Recombinants, double crossover AC, CB
8	a	C	b	2		

Gene Order

The most likely gene order minimizes the sum of pairwise recombination frequencies within a three-gene interval, which would be:

- 38.4 for **ABC** (28.4% for **AB** + 10% for **CB**)
- 30.8 for **ACB** (20.8% for **AC** + 10% for **CB**)
- 49.2 for **CAB** (20.8% for **AC** + 28.4% for **AB**)

Thus, the most likely gene order is **ACB**. In other words, the interval between **A** and **B** can be subdivided into the intervals between **AC** and **CB**.

Class	Genotype of gamete from heterozygous parent			Number		Origins
1	A	C	B	179	}	Parentals, no crossover
2	a	c	b	173		
3	A	c	b	52	}	Recombinants, single crossover AC
4	a	C	B	46		
5	A	C	b	22	}	Recombinants, single crossover CB
6	a	c	B	22		
7	A	c	B	4	}	Recombinants, double crossover AC, CB
8	a	C	b	2		

Expressed yet another way: incorrectly ordered genes would increase the total map length because part of the recombination events would be counted twice. If **ACB** is the true order, then the genetic length of, e.g., **ABC** would be inflated, because recombinants for the segment **BC** would be counted two times: for the interval **BC**, in addition to the same interval within the segment **A**(**C**)**B**. Algorithms of mapping programs use this principle (minimizing the genetic distance) for three-point analyses.

Double Crossovers

The two-point recombination frequency between **A** and **B** (28.4%) differs from the sum of recombination frequencies for **AC** and **CB** (30.8%). The reason for this discrepancy is the occurrence of double crossover events. These are two crossovers in a single meiosis within an interval of interest and the second crossover reverses the effect of the first crossover, i.e., the second crossover returns the **B** allele to the original position before the first crossover. For this reason, by only taking recombinants between **A** and **B** into consideration, double crossovers cannot be observed. Because a double crossover exchanges chromosome segments within an interval of two genes, the linkage phase (coupling) of those two genes remains unchanged.

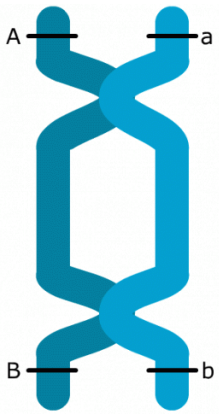


Fig. 13 Double crossover event.

Observing Double Crossovers

Only by adding a gene like **C** in between **A** and **B**, it is possible to observe double crossovers. In the case of the interval between genes **A** and **B**, six double crossovers were observed. In consequence, recombination and crossover frequencies are not identical. The larger the genetic interval, the larger the discrepancy between recombination and crossover frequencies, because even-numbered crossover events within a pair of genes go undetected. By adding an additional gene in this interval, at least some double crossovers can be detected. This leads to detection of additional recombination events. For this reason, the recombination frequency between **A** and **B** in the example is increased, after adding **C** in between those two genes, because 6 double crossovers (= 12 additional recombination events) could be detected. Those 12 additional detectable recombination events explain for the 2.4% difference between recombination frequencies detected for the gene pair **A** and **B** with or without inclusion of **C**.

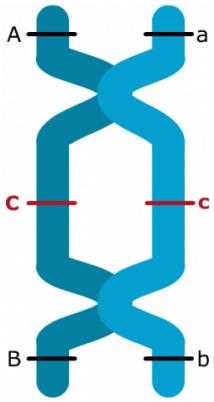


Fig. 14 Double crossover event with recombination.

Phase Analysis

Recombinants resulting from double crossovers are always in the lowest frequency (class 7 and 8, respectively in this table). To determine which allele is in the middle, a convenient method is to find out which allele in the double crossover recombinants has changed its linkage phase with the other parental alleles (in classes 7 and 8, allele **C**/**c** has changed its linkage phase with the other alleles).

Class	**Genotype of gamete from heterozygous parent**			**Number**		**Origins**
1	A	C	B	179	}	Parentals, no crossover
2	a	c	b	173		
3	A	c	b	52	}	Recombinants, single crossover AC
4	a	C	B	46		
5	A	C	b	22	}	Recombinants, single crossover CB
6	a	c	B	22		
7	A	c	B	4	}	Recombinants, double crossover AC, CB
8	a	C	b	2		

Coefficient of Coincidence and Interference

Crossover events in adjacent chromosome regions might affect each other, a phenomenon called **interference**. Most typically, a crossover event in one region tends to suppress a crossover in the adjacent regions. The extent of interference is expressed by the **coefficient of coincidence**, which is equal to the observed frequency of double crossovers / expected frequency of double crossovers.

Class	**Genotype of gamete from heterozygous parent**			**Number**		**Origins**
1	A	C	B	179	}	Parentals, no crossover
2	a	c	b	173		
3	A	c	b	52	}	Recombinants, single crossover AC
4	a	C	B	46		
5	A	C	b	22	}	Recombinants, single crossover CB
6	a	c	B	22		
7	A	c	B	4	}	Recombinants, double crossover AC, CB
8	a	C	b	2		

The expected frequency of double crossovers is the product of two single crossovers in adjacent regions assuming there is no interference.

In this example, this expected frequency is 0.21 (recombination frequency for AC) * 0.10 (recombination frequency for CB) = 0.021.

The observed frequency of double crossover events is 6/500 in the example, resulting in 0.012. Thus, the **coefficient of coincidence** in this example is 0.012/0.021 = 0.58.

Interference is defined as 1 – coefficient of coincidence, which would be 0.42 in this example. A value of zero for interference would mean that a crossover in one region does not affect crossovers in the adjacent region. Interference of 1 means, that crossovers in one region suppress crossovers in the adjacent region. Negative values are possible and have been reported in some instances, which means that crossovers in one region stimulate crossovers in the adjacent region.

Map Functions

Measurement Units

The purpose of genetic maps is to report the length of chromosome intervals, chromosomes, and whole genomes. Since recombination frequencies converge to a value of 50% as reported above, indicating the absence of linkage, recombination frequencies are not additive and, thus, not useful to describe the distance between genes that are located far apart. When recombination frequency reaches 50%, it would be impossible to tell whether the genes are located far apart on the same chromosome or on different chromosomes.

Instead, estimates of the number of crossover events are used as an additive measure of genetic map distances. The unit for measuring genetic distances is Morgan (M), or usually **centiMorgan** (**cM**). In contrast to recombination frequencies, map units expressed in cM are additive. One Morgan reflects the observation of one crossover event per single meiosis. One cM is a distance between genes that produces 1% recombinants in the offspring. Typical lengths of genetic maps in maize, for example, vary between 1,600 to 2,000 cM, which means that on average, 1.6 – 2 crossovers occur per chromosome and single meiosis in maize (maize has 10 homologous chromosome pairs).

Frequency Conversion

As mentioned above, direct observation of crossover events is cumbersome. For that reason, most genetic maps published to date are based on the conversion of recombination frequencies into crossover frequencies. The main obstacle to translating recombination frequencies into crossover frequencies is the variable and unknown degree of interference in different genome regions. While it has been possible in the earlier example to determine the degree of interference, and thus frequency, of double crossover events, in the genetic interval between **A** and **B** by adding **C**, the degree of interference between **AC** and **CB** is unknown. This could be addressed by observing segregation of further genes within these two regions (if available), but this issue could ultimately only be addressed by complete genome sequencing of all offspring in a mapping population, which at this point is still too costly.

Visual Relationship

Instead, **map functions** have been developed, that translate recombination frequencies into

crossover frequencies, and thus cM (see Fig. 14 below). Figure 14 clearly shows, that there is an approximately linear relationship between recombination rates (y-axis) and crossover rates (x-axis). However, with increasing map distances, recombination rates converge to 50%. In other words, gene pairs with crossover rates of 80 cM or 200 cM, respectively, would be nearly indistinguishable based on recombination rates, which would result in recombination rates between 40 and 50%.

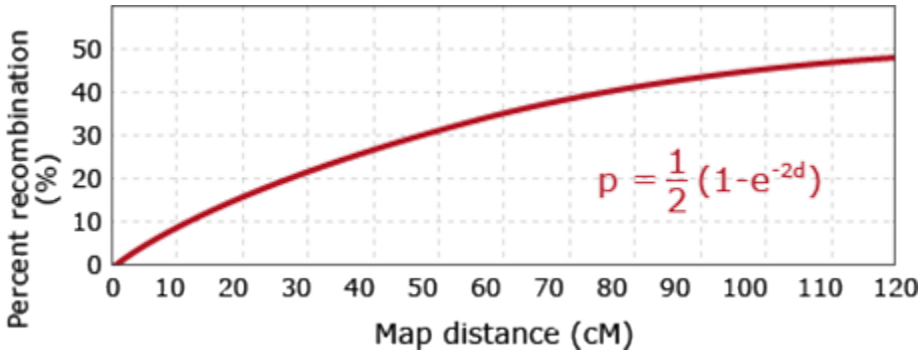


Fig. 15 Relationship between crossover rates (x-axis) and recombination rates (y-axis). Adapted from Russell, 2010.

The various available map functions make different assumptions on the extent of interference. For example, the Haldane mapping function assumes absence of interference. In contrast, the Kosambi function assumes the presence of interference.

Other Types of Maps

Genetic maps can be generated in other ways than using testcrosses. Examples include somatic cell hybridization and tetrad analysis. In plants, interspecies addition lines such as oat-maize addition lines created by distant hybridization have been developed as tool for mapping of genes. If two genes appear on the same addition segment, they are genetically linked. Besides genetic maps, cytological and physical maps can be established.

Cytological maps show gene orders along each chromosome as determined by cytological methods whereas physical maps are measured in base pairs as determined by DNA sequencing. With rapid progress in sequencing technology and an increasing number of sequenced plant genomes, physical maps gain in importance. Complex plant genomes like the maize genome are billions of base pairs long. When comparing genetic and physical maps, the order of genes is conserved. However, the relative distances between genetic and physical maps might vary substantially. The reason is that crossover events are not evenly distributed in genomes. Usually, crossover events tend to be suppressed in centromere and repetitive DNA regions, whereas they are enhanced in gene-rich regions.

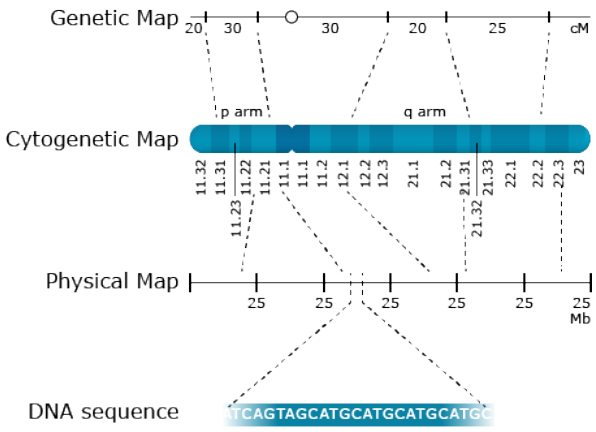


Fig. 16 A genetic map is a type of chromosome map that shows relative locations of genes and other important features. Adapted from NIH-NHGRI 2011.

Cytogenic Map

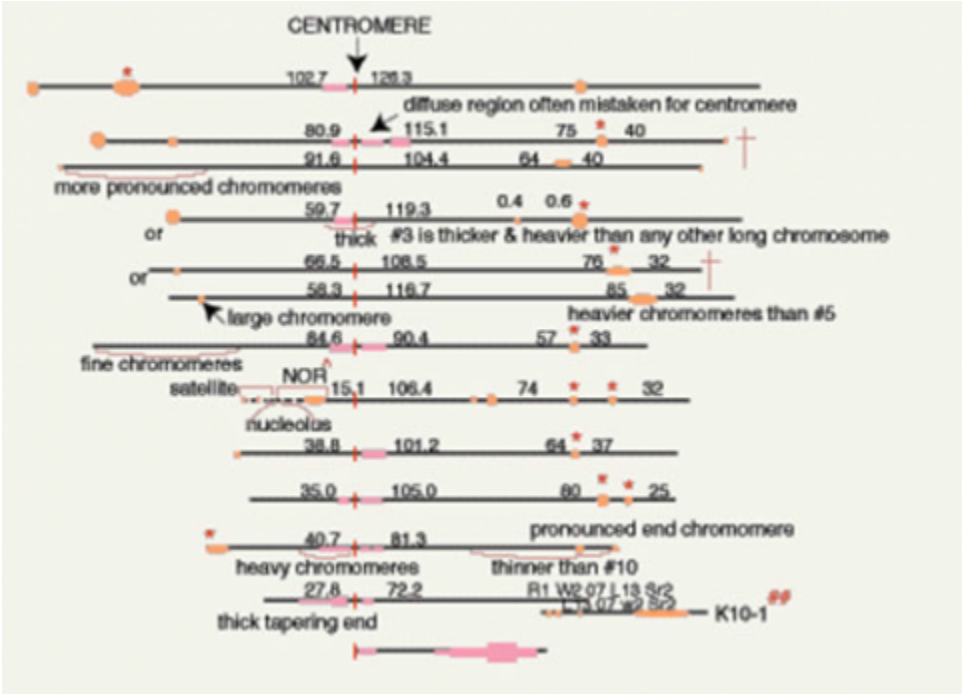


Fig. 17 Cytogenetic map of maize chromosomes. Illustration by Neuffer et al., 1997. Used with permission.

Factors Influencing Linkage Mapping

Linkage mapping based on testcrosses can be affected by selection or incomplete penetrance, among others. Selection in the most extreme case would be due to lethality of gametes (gametic selection) or zygotes (zygotic selection). If a backcross is used for linkage detection, as described above, lethality of male gametes carrying for example the *a* allele would lead to only two classes of BC progeny, if **AaBb** is crossed as pollinator to **aabb**. In that case, only **AaBb** (parental) and **Aabb** (recombinant) genotypes would be obtained. Zygotic selection affects the viability of particular genotypes. If the **aa** genotype in the example above is lethal, then the **aa** offspring derived from self-pollination of an **AaBb** genotype would be missing. Incomplete penetrance means that a genotype that is supposed to express, for example, red flowers, has to a certain extent white flowers. In other words, there is no 100% match between genotype and phenotype, but due to environmental factors, the phenotype might differ. As for selection, incomplete penetrance alters the frequency of expected genotypes in testcrosses, which is the basis for detection of linkage.

Consequences and Applications of Linkage

The main application of linkage is in genetic mapping of genes using molecular markers. Once genes have been mapped and closely-linked markers identified, those markers can be used for marker-aided selection procedures. Technological progress in DNA methods has been and still is rapid, so that thousands of markers can be produced at low cost in any species of interest. Moreover, novel genomic selection strategies addressing complex inherited traits are being developed.

Linkage can in some cases be confused with pleiotropy. If a favorable character (e.g. resistance) is always inherited together with an unfavorable trait (e.g., lodging), a negative pleiotropic effect might be assumed, which might alternatively be caused by two closely linked genes. Whereas close linkage can be resolved to find favorable genotypes for both traits, this is not true for pleiotropy. Linkage reduces the possible genetic variation in small populations. With increasing numbers of generations, or population sizes, genetic variation can be increased. Similarly, inbreeding reduces the opportunity for effective recombination.

Linkage Disequilibrium

Genotype Distribution

Although allele frequencies at individual loci are expected to be stable in the case of random mating, genotype frequencies at two or more loci jointly do not achieve this equilibrium after one generation of random mating.

To illustrate this point, consider two populations, one consisting of entirely **AABB** genotypes and the other consisting entirely of **aabb** genotypes. Assumed they are mixed equally and allowed to randomly mate. The first generation would consist of the three genotypes **AABB**, **AaBb**, and **aabb** in the proportions 1/4 : 1/2 : 1/4. However, for two loci, each with two alleles, nine genotypes are possible. (For n alleles at each locus and k loci, there are: $(\frac{n(n+1)}{2})^k$ possible genotypes). Continued random mating would produce the missing genotypes, but they would not appear at the equilibrium frequencies immediately.

Equilibrium

Consider the following examples based on two alleles at each of two loci:

- **Alleles**: A a B b
  - **Allele frequencies**: P_A P_a P_B P_b

- **Gametic Types**: AB Ab aB ab
  - **Gametic Frequencies**: P_{AB} P_{Ab} P_{aB} P_{ab}

In linkage equilibrium, the expected gamete frequencies can be calculated from the marginal allele frequencies. For example, in equilibrium, the frequency of gamete **AB** (P_{AB}) would be expected to be equal to the product of the frequencies of the **A** allele (P_A) and the **B** allele (P_B).

This is valid under the following conditions: $P_A + P_a = 1$; $P_B + P_b = 1$; and $P_{AB} + P_{Ab} + P_{aB} + P_{ab} = 1$.

If, for example, the allele frequencies of $P_A = P_a$ and $P_B = P_b$ are 0.5, then the frequencies of all gametes are 0.25.

A measure for Disequilibrium, $D = P_{AB} - P_A{}^*P_B$. $D = 0$ in the case of equilibrium. If D differs from 0, it reflects presence of Disequilibrium. In other words, the frequency of a gamete differs from its expected frequency based on marginal probabilities of the respective individual alleles.

LD and Mapping

Linkage disequilibrium is the non-random association of alleles at different loci. LD is extensively used in mapping human disease genes using natural populations (Association mapping).

In plants, gene mapping has been conducted mainly by using mapping families because of the ease with which mapping families are created, but LD mapping using natural populations is increasing rapidly because such populations are large in size and have much greater allelic diversity.

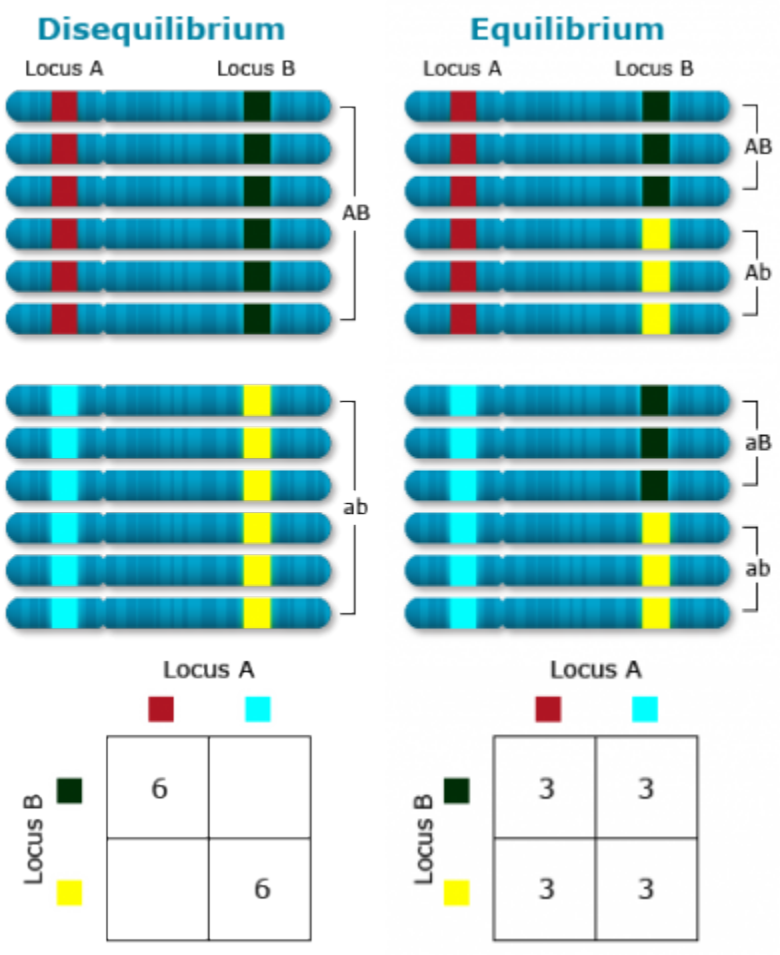


Fig. 18 Linkage disequilibrium and equilibrium. When LD is present, all individuals possessing red alleles in locus A have green alleles in locus B. When the two loci are in equilibrium, individuals having red alleles in locus A could have any alleles in locus B. Adapted from Rafalski, 2002.

LD Statistic D'

$$|\mathbf{D'}| = \frac{D_{AB}^2}{min(p_A p_b, p_A p_b)}$$

for DAB < 0

$$|\mathbf{D'}| = \frac{D_{AB}^2}{min(p_A p_b, p_A p_B)}$$

for DAB > 0

Dissipation

It can be shown that after t generations of random mating, the remaining disequilibrium is given by:

$$D_t = D_0(1-c)^t$$

where, D_0 is the disequilibrium in generation 0 and c is the recombination fraction, with c = 0.5 for independently segregating loci, which is identical to a recombination frequency of 50% (the range of c is from 0 to 0.5, whereas the range of r is from 0% to 50%). The dissipation of disequilibrium relative to generation 0 is given in Fig. 18.

Recombination and LD

Generally, deviations from independence at multiple loci are referred to as linkage disequilibrium, even if genetic linkage is not the cause (in other words, alleles are not physically linked). Unless two loci are known to reside on the same chromosome, the term Gametic Disequilibrium should be used to describe disequilibrium among loci. Whereas recombination and crossover frequencies, as mentioned initially, are used to describe the distance between genes from a chromosomal perspective, linkage disequilibrium is mostly used to describe a property of populations. However, both terms are closely related.

Genetic Markers

Overview

Genetic variation results from differences in DNA sequences and, within a population, occurs when there is more than one allele present at a given locus. Such populations are referred to as populations that are **polymorphic** or segregating at that locus. The opposite situation is when all members of the population are homozygous for the same allele, in which case the population is said to be **fixed** or **monomorphic** for that allele. A **genetic marker** is a DNA sequence that exhibits polymorphism among individuals and can thus be used to identify a particular locus (although not necessarily a gene) on a particular chromosome; the marker itself may be part of a gene or may have no known function. Markers are inherited in a Mendelian fashion and facilitate the study of inheritance of a trait or sometimes a linked gene. Markers are used to identify, **map**, and isolate genes, select desired genotypes, and detect genetic variation

or determine genetic relationships among individuals. Markers are regions of genomes that are heritable, often easy to document, and useful for detecting genetic variation.

Three Types

Genetic markers generally do not represent target genes of interest to a breeding program, but instead are useful as 'signs' or 'tags', particularly when they are closely linked to genes that control a trait of interest. A genetic map constructed with genetic markers is similar to a road map. Linkage groups in a genetic map represent roads whereas individual markers on each linkage group represent signs or landmarks that help plant breeders to navigate through the plant genome and find the genes of interest.

There are three major categories of markers.

- **Morphological markers**
- **Biochemical markers**
- **Molecular markers**

Morphological Markers

These types of markers (also called **visible** or **classical** markers) are phenotypic traits with only a few distinct **morphs** or variants (e.g., flower color or seed shape), usually due to one or perhaps two gene loci so they are not strongly affected by the environment. Inheritance patterns of visible and morphological characters have been used to map genes to particular chromosome segments and to identify linkage groups. Such markers are limited in number compared to the abundance of DNA markers, however, and may be influenced by developmental stage of the plant.

Biological Markers

Isozymes (sometimes called allozymes) are allelic variants of a single enzyme that share the same function, but may differ in level of activity due to differences in amino acid sequence. Isozymes are proteins for which variation can be detected by differential separation using **electrophoresis**, a technique for separating macromolecules (DNA, RNA, protein) on a gel by means of an electric field and specific chemical staining.

Isozymes have **codominant** expression, meaning that both homozygotes can be distinguished from the heterozygote and neither allele is recessive. In contrast to codominant markers, **dominant** markers are either present or absent.

In comparison to visible **polymorphisms,** they reveal more of the underlying genetic variation. However isozymes are gene products, so they reveal only a small subset of the actual variation in DNA sequences between individuals and do not reveal variation in the non-coding regions of the genome. In general, such markers are limited in number and have limited use in genetic mapping studies.

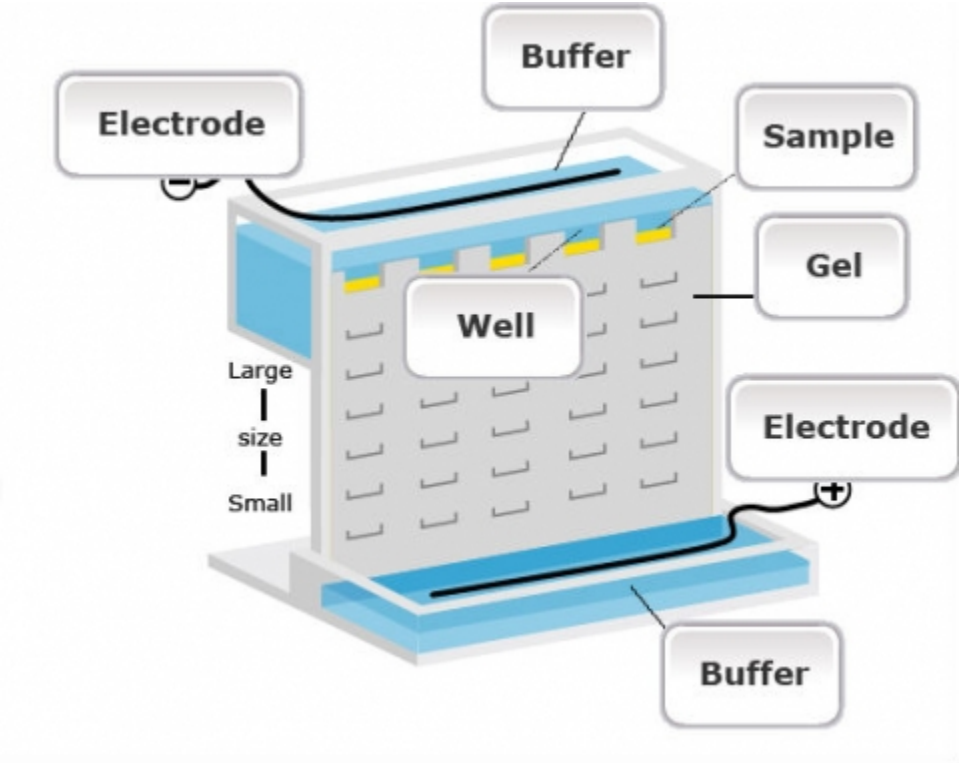


Fig. 20 Electrophoresis is a laboratory technique used to separate DNA, RNA, or protein molecules based on their size and electrical charge. Illustration adapted from NIH-NHGRI, 2011.

Molecular Markers

Molecular or DNA markers reveal sites of variation in DNA. Variability in DNA facilitates finer scale **mapping** and detection. Mapping is the process of making a representative diagram cataloging genes and other features of a chromosome and showing their relative positions. Many of these molecular markers avoid the limitations associated with visible and biochemical

markers. They facilitate the evaluation of genome-wide coverage and are not affected by environmental factors or developmental stages. They allow high resolution of genetic diversity to be detected. Molecular markers have added substantial amounts of information to our genetic maps.

FYI: Molecular Markers

Any DNA sequence can be genetically mapped, like genes leading to plant phenotypes as long as there is a polymorphism available for the sequence to be mapped, i.e., two or more different alleles. This can basically be a **single nucleotide polymorphism** (**SNP**), a single nucleotide variant at a particular position within the target sequence, or an insertion/deletion (**INDEL**) polymorphism. Any target sequence can be amplified by the **Polymerase chain reaction** (**PCR**), and subsequently be visualized to generate "molecular phenotypes" comparable to visual phenotypes, that can be observed by using appropriate equipment.

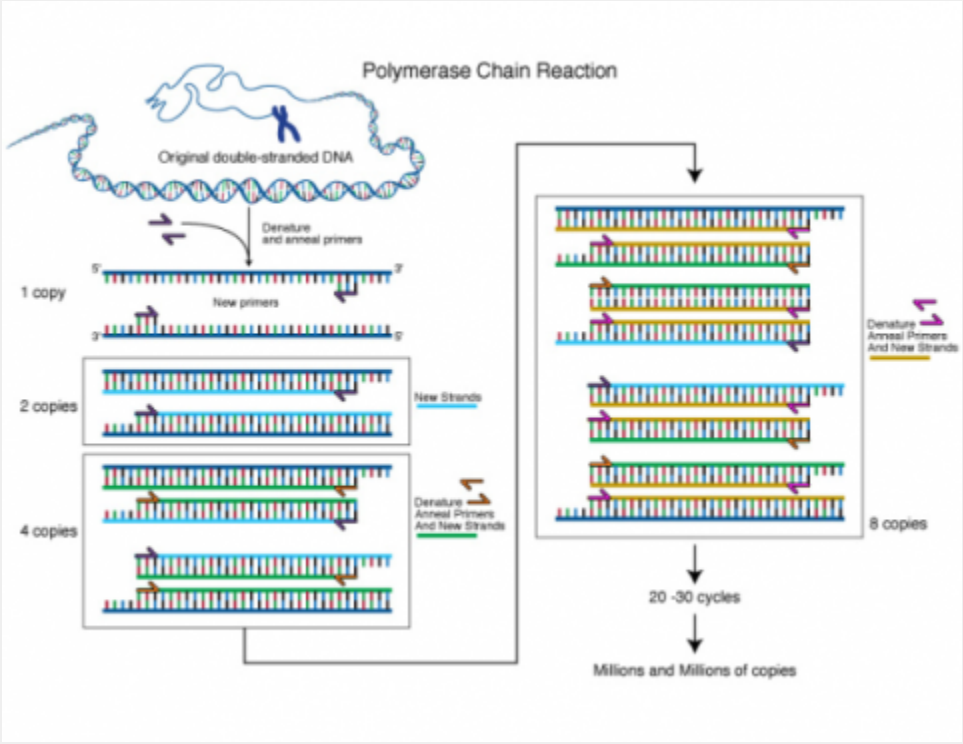


Fig. 21 A depiction of Polymerase chain reaction (PCR). Illustration by NIH-NHGRI, 2011.

Various molecular methods have been developed to visualize SNPs or INDEL polymorphisms at low cost and high throughput, which will be presented in detail in the Molecular Genetics and Biotechnology course. The main use of those SNPs and INDEL polymorphisms is as molecular markers. By genetic mapping as described above, linkage between genes affecting agronomic traits or morphological characters, and DNA-based SNP or INDEL markers can be established. It can be more effective in the context of plant breeding, to select indirectly for such DNA markers, than directly for target genes. This is due to lower costs for DNA analyses, the ability to run multiple such assays (for multiple target genes) in parallel, the ability to select early and to discard undesirable genotypes or to perform selection before flowering, codominant inheritance of markers, among others.For example, both ginkgo trees (*Ginkgo biloba*) and asparagus (*Asparagus officinalis*) are dioecious species. Male plants are preferred for ginkgo tree because fruits produced from female trees have an unpleasant smell whereas male asparagus plants are preferred because of their higher yield potential. Unfortunately, sex expression will take years to occur for both species. If a DNA marker that either directly affects sex expression or is linked to genes that affect sex expression can be identified, selection of male plants can be conducted in early seedling stage rather than waiting for many years. Occurrence of environmental conditions favoring selection for disease, insect-resistant plants or drought-tolerant plants such as the prevalence of the particular disease or insect or drought is not always reliable. Selection using DNA markers can overcome these limitations as they are not affected by the environment.

Polymorphism

Polymorphism involves one of two or more variants of a particular DNA sequence. The most common type of polymorphism involves variation at a single base pair, also called single nucleotide polymorphism (SNP) (Fig. 22). Polymorphisms can also be much larger in size and involve long stretches of DNA. Tandem repeat is a sequence of two or more DNA base pairs that is repeated in such a way that the repeats are generally associated with non-coding DNA. In contrast, SNPs can sometimes be identified that occur within coding sequences (that is within genes), as well as in non-coding DNA.

Fig. 22 Example for a SNP (yellow highlighted) in a population of six diploid genotypes. Individuals 1, 4, 5, 6 are T/C heterozygotes, individual 2 a C/C homozygote, and individual 3 a T/T homozygote. Adapted from NIH-NHGRI, 2011.

Types of Biochemical/Molecular Markers

There are a variety of biochemical and molecular markers available. Table 2 on the next page summarizes features of a number of the common ones:

- RFLP — Restriction Fragment Length Polymorphisms
- RAPD — Random Amplified Polymorphic DNA
- AFLP — Amplified Fragment Length Polymorphisms
- SSR — Simple Sequence Repeats (also known as **microsatellites**)
- SNP — Single Nucleotide Polymorphisms
- VNTR — Variable Number of Tandem Repeats

Widely-Used Markers

Table 2. Comparison among widely used molecular markers. Adapted from Nageswara-Rao and Soneji, 2008.

	Isozymes	RFLP	RAPD	AFLP	SSR	SNP
Protein- or DNA-based	Protein-based	DNA-based	DNA-based	DNA-based	DNA-based	DNA-based
No. of loci	30-50	100s	~Unlimited	~Unlimited	10s	10s
Degree of polymorphism	Low-medium	Meduim-high	Medium-high	Medium-high	High	High
Nature of gene action	Codominant	Codominant	Dominant	Dominant	Codominant	Codominant
Reproducibility	High	High	Low-medium	Medium-high	High	High
Amount of DNA per sample	Not applicable	mg	ng	ng	ng	ng
Method*	Biochemical	DNA-DNA hybridization	PCR	PCR	PCR	PCR
Ease of array?	Easy	Difficult	Easy	Moderate	Easy-moderate	Easy
Can be automated?	Difficult	Difficult	Yes	Yes	Yes	Yes
Equipment cost	Inexpensive	Expensive	Moderate	Expensive	Expensive	Expensive
Development cost	Inexpensive	Expensive	Moderate	Expensive	Very	Expensive
Assay cost	Inexpensive	Expensive	Moderate	Expensive	expensive	Expensive

* 'PCR' means Polymerase Chain Reaction amplification of genomic DNA fragments, a method that uses short, single-stranded DNA sequences, known as primers, to hybridize with the sample DNA Table 2. Comparison among widely used molecular markers. Adapted from Nageswara-Rao and Soneji, 2008.

SSR and SNP Markers

SSR markers remain useful to plant breeders due to their abundance and convenience with which they are assessed, but they serve most likely as linked markers. SNP markers, however

can either be linked to or directly reside in a gene of interest and are hugely abundant. For these reasons, they are increasingly becoming the marker of choice.

Uses of Molecular Markers

Molecular markers are useful for both applied and basic genetic research. Here are some examples:

Indirect selection criteria in breeding programs (marker-assisted selection)

This is one of the most important and widely used molecular techniques in applied plant breeding programs today. RFLPs, SSRs, and SNPs enable breeders to indirectly select for a desired trait. Ordinarily, the DNA sequence of a molecular marker does not itself code the gene for the trait, but rather, its presence is correlated or linked with the gene for the particular trait. Thus, the breeder can indirectly select for the trait by directly selecting for the molecular marker—the DNA fragment and the gene encoding the trait are linked. The closer their physical proximity on a chromosome, the greater the probability that they will remain linked and not be separated through a recombination event in subsequent generations. As long as the gene for the trait and the marker remain linked, the marker is a useful selection criterion. Ultimately, however, potential lines still must be field-tested to verify the expression of the desired phenotype.

Identify quantitative trait loci (QTLs)

Molecular markers linked to genes contributing to the expression of polygenic or quantitative traits can be used to more efficiently identify and select individuals possessing the genes. It is more difficult to use conventional breeding approaches to identify plants that have accumulated the genes necessary to obtain the desired quantitative trait.

Genetic mapping

Molecular markers provide a means to map genes to more specific chromosome segments than is possible using visible markers.

Determine genetic relationships

The more molecular markers individuals have in common, the more closely related they are.

Genetic Diversity and Conservation

Genetic relationships within families, genera, species, or cultivars can be determined from molecular markers. Much information about the evolution of crops has been learned using molecular markers. The markers also enable breeders to monitor the genetic diversity among breeding lines to broaden the genetic base and reduce the risk of widespread genetic vulnerability to detrimental conditions.

Molecular Marker 'Fingerprints'

Individuals possessing more markers in common than could occur by random chance are closely related. Such molecular fingerprints have been used successfully in court to prove the misappropriation of proprietary breeding lines.

Isolate genes

Molecular markers are used to map candidate genes on a much finer scale and can eventually isolate candidate genes by positional cloning. Isolated genes can be used to study gene regulation or to directly improve agronomic performance by genetic transformation. Although molecular markers have many applications and provide useful tools to plant breeders, lines must still be evaluated under normal production conditions before their release.

References

Falconer, D.S., and Trudy F.C. Mackay. Introduction to Quantitative Genetics (4th edition). San Francisco, CA: Benjamin Cummings.

Fehr, Walter R. 1987: Principles of Cultivar Development. Macmillan Publishing Company: New York.

Nageswara-Rao, M., J.R. Soneji, C. Chen, S. Huang, and F.G. Gmitter. 2008. Characterization of

zygotic and nucellar seedlings from sour orange-like citrus rootstock candidates using RAPD and EST-SSR markers. Tree Genet Genomes. 4:113-124. DOI: 10.1007/s11295-007-0092-2

Neuffer M.G., Coe, E.H. and Wessler, S.R. (1997) Mutants of Maize. Plainview, NY: Cold Spring Harbor Laboratory Press

NIH-NHGRI (National Institutes of Health. National Human Genome Research Institute). 2011. Talking Glossary of Genetic Terms. [available online September 23, 2011, http://www.genome.gov/glossary/.

Neuffer M.G., Coe, E.H. and Wessler, S.R. (1997) Mutants of Maize. Plainview, NY: Cold Spring Harbor Laboratory Press

Pierce, Benjamin A. 2008: Genetics: A Conceptual Approach. W.H. Freeman and Company: 160-199, 335-339. New York, NY.

Rafalski, A. 2002 Applications of single nucleotide polymorphisms in crop genetics. Current Opinion in Plant Biology 2002, 5:94–100

Russell, Peter J. 2012: iGenetics: A Molecular Approach. Pearson Education, Inc., San Francisco, Calif.

How to cite this chapter: Lübberstedt, T., A. Campbell, D. Muenchrath, L. Merrick, and S. Fei. 2023. Linkage. In W. P. Suza, & K. R. Lamkey (Eds.), *Crop Genetics*. Iowa State University Digital Press. DOI: 10.31274/isudp.2023.130

Population Genetics

Laura Merrick; Kendra Meade; Arden Campbell; Deborah Muenchrath; and William Beavis

Introduction

Population genetics is a sub-discipline of genetics that characterizes the structure of breeding populations. The forces of mutation, migration, selection and **genetic drift** will alter the structure of populations. In this introductory module we will focus on characterizing population structure at a single locus. In more advanced modules you will learn how to characterize populations based on the multi-dimensional space determined by multiple loci throughout the genome.

Fig. 1 A bed of tulips at Pecherskyi Landscape Park in Ukraine. Photo by OZ-OK. Licensed CC-BY-SA 3.0 via Wikimedia Commons.

Learning Objectives

- Understand the importance of a reference population.
- Become familiar with modeling and estimation of genetic variation.
- Understand the principles of allele frequency, genotype frequency, and genetic equilibrium in populations.
- Be aware of the conditions required for Hardy-Weinberg Equilibrium (HWE).
- Examine the forces that cause deviations from HWE.

Two possible challenges are described in the following scenarios:

Scenario 1—Fate of a Transgene

Fig. 2 Maize and bean plants in a field in Quiché, Guatemala. Photo by Fabian Hanneforth. CC-SA 3.0 via Wikimedia Commons.

Imagine a community of small farms in a valley located in the highlands of Central America. The farmers of this community produce grain from an open-pollinated maize variety that is adapted to their preferred cultural practices. They also select partial ears from about 5% of their better performing plants to be used for seed in their next growing season.

One day, a truck filled with seed of a transgenic insect-resistant hybrid overturns on the highway while passing through the valley. 99.999% of the seed is recovered, but about 500 kernels remain in a farmer's 10-acre field adjacent to the highway. The transgenic seeds germinate and grow to maturity alongside the planted open-pollinated variety. You are asked to determine the fate of an insect-resistant transgene in this valley.

Scenario 2—Fixation of an Allele

Imagine a naturally occurring allele at a locus that regulates the structure of carbohydrates in the wheat kernel; with the allele the carbohydrates in the kernel have low glycemic indices. For the last 100 years hard-red winter wheat varieties have not been selected for low glycemic indices, but with the emergence of a Type II diabetes epidemic, there is a demand for low glycemic carbohydrates in hard-red winter wheat varieties. How will you develop a breeding population in which this allele is fixed, that is the frequency of this allele = 1.0?

Fig. 3 Hard red winter wheat kernels. Photo by U.S. Department of Agriculture.

Fields of Genetics

Fig. 4 **Transmission genetics** studies how traits are passed from an individual to its progeny

These challenges are fundamentally about population genetics. In this section, you have the opportunity to successfully address these types of challenges by learning how to model and estimate allelic frequencies and the forces that affect population structures. In the study of population genetics, the focus shifts away from the individual (which is the focus for **transmission genetics**) and the cell (which is the focus for **molecular genetics**) to emphasis on a large group of individuals—a Mendelian population—that is defined as a group of interbreeding individuals who share a common set of genes.

This module will include a discussion of **inbreeding**, which is one type of mating of individuals that is often of particular significance to plant breeders.

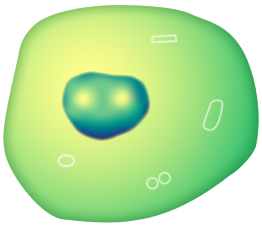


Fig. 5 **Molecular genetics** focuses on DNA operating within the cell.

Inbreeding is the mating of individuals that are more closely related than individuals mated at random in a population. Self-pollination (mating of an individual to itself) represents the most extreme form of inbreeding.

Fig. 6 **Population genetics** studies the forces that affect groups of interbreeding individuals who share a common set of genes.

Reference Population

Goals

Population genetics has three major goals, all of which are interrelated (Conner and Hartl, 2004):

- Explain the origin and maintenance of genetic variation.
- Describe the genetic structure of populations, i.e., the patterns and organization of genetic variation.
- Recognize the mechanisms that cause changes in allele and genotypic frequencies.

Fig. 7 Analysis of rice crop populations. Photo by Iowa State University.

Similar to **quantitative genetics**, population genetics is concerned with application of Mendelian principles and is amenable to mathematical treatment. Understanding population genetics will require you to apply concepts from high school algebra.

Description

In order to understand the genetic structure of a population, it is necessary to establish a standard reference population so that the breeding population can be characterized relative to the standard.

Consider an 'ideal' population that is infinitely large. Further consider development of sub-populations as in Figure 6, described in Falconer and Mackay (1996).

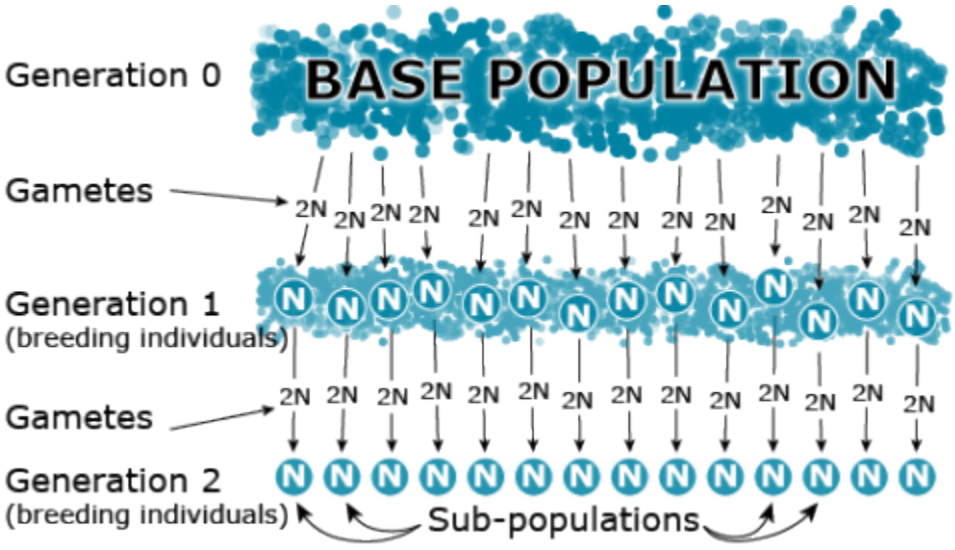


Fig. 8 Diagram of the subdivision of a single large base population into a number of sub-populations.

Note that the sub-populations depicted in the figure above are based on a genetic sampling process that is affected by reproductive biology of the species. The reproductive mode of most plant species can be classified as **sexual** or **asexual** Species that reproduce sexually are generally categorized into three types of mating systems — primarily cross-pollinated, primarily self-pollinated, or a mixture of self- and cross-pollinated. Asexual modes of reproduction include three main categories: **vegetative** or clonal propagation, and **apomixis**. Under different mating systems (e.g., random vs. inbreeding) different genotypic frequencies will be generated from the same allele frequencies. With sexually reproducing individuals, mating combines alleles in the pool of haploid gametes produced by meiosis into genotypes in the diploid individuals.

In the ideal **model population** depicted in Fig. 8, we make the following assumptions:

- The base population is extremely large (too large to count)
- No migration between sub-populations
- Non-overlapping generations
- Number of breeding individuals is the same in each sub-population
- Random mating within a sub-population
- No selection
- No mutation

Models such as that shown above are theoretical abstractions. Models provide methods to simulate real-life situations and they are used for two principal reasons: 1) to reduce complexity, allowing underlying patterns to become more visible and 2) to make specific predictions to test with experiments or observations (Connor and Hartl 2004).

Discussion

Discuss the two challenges described earlier with respect to each reference population:

Photo by the U.S. Department of Agriculture.

For **Scenario 1—Fate of a Transgene**, characterize the breeding population. Assume that there are 100 10-acre farms in the Central American valley, where farmers plant about 10,000 maize kernels per acre.

For **Scenario 2—Fixation of an Allele**, determine how many hard red winter wheat varieties exist for the Southern Great Plains region. The number can include all historical varieties grown in the region. Assume that you have identified one additional ancient accession of hard red winter wheat that has the desirable allele for low glycemic carbohydrates. Assume that these varieties represent the lines you will use for your basic breeding population. Characterize this breeding population.

Photo by Fabian Hanneforth. Licensed under CC-SA 3.0 via Wikimedia Commons.

Allele and Genotypic Frequencies

Model

We first model a single locus with only two alleles (e.g., presence or absence of a transgene) in an ideal breeding population of diploid individuals. Define the following:

- **N** = Number of breeding individuals in a sub-population (population size)
- **t** = Time in generations with base population at t_0
- **q** = Frequency of a particular allele at a locus within a sub-population
- **p** = **1** − **q** = Frequency of other allele at a locus within a sub-population
- $\bar{p}$ = Frequency in the whole population (the mean of p)
- $\mathbf{p_0}$ = Frequency of p in the base population
- $\mathbf{q_0}$ = Frequency of q in the base population

Because of the assumptions associated with an ideal reference population, $\overline{q}$ = q_0 at any stage or generation of the sampling process, so q_0 can be used interchangeably with $\overline{q}$.

Equations

The alleles, allele frequencies, genotypes and genotypic frequencies can be represented as follows:

	Alleles		**Genotypes**		
	A	a	AA	Aa	aa
Frequency	p	q	P_{AA}	P_{Aa}	P_{aa}

Where,

$$p + q = 1$$

and

$$P_{AA} + P_{Aa} + P_{aa} = 1$$

The relationship between allele frequencies and genotype frequencies can be expressed as follows:

$$p = P_{AA} + \frac{1}{2}P_{Aa}$$

and

$$q = P_{aa} + \frac{1}{2}P_{Aa}$$

Hardy-Weinberg Equilibrium

Concept of Genetic Equilibrium

Plant breeders recombine and select the alleles present in the **gene pool**. The gene pool of a population is the total of all alleles within a population, and consists of all of the genes shared by individuals in the population. Gene pools are described in terms of allele and genotype frequencies. Knowing the frequency with which desired (or undesirable) alleles occur in the gene pool of the population influences the choice of breeding population(s), breeding method, and likelihood of progress. The breeding population must contain not only sufficient genetic variability to allow selection, but also have favorable alleles present in high enough frequencies to facilitate their selection and allow efficient breeding progress to occur.

- **Allele frequency** (often also called gene frequency) — the proportion of contrasting alleles present in the gene pool of a population.
- **Genotype frequency** — the proportion of various genotypes present in a population.

Assumptions

The frequencies of specific alleles and genotypes in a large, **random mating** population will reach equilibrium and will remain in equilibrium with continued random mating. This tendency toward equilibrium is the foundation of a model called the Hardy-Weinberg Law or Hardy-Weinberg Equilibrium (HWE). This law states that

The probability of two alleles uniting in a zygote is the product of the frequency of the alleles in the population

The law makes several assumptions.

- There are two alleles at a gene locus.
- The population is large (that is, the number of breeding individuals is in the hundreds, rather than in the tens).
- The population is random-mating.

Frequencies

Hardy-Weinberg Equilibrium mathematically describes the relationship between allele frequency and genotype frequency. According to the Hardy-Weinberg law, if the frequencies of two contrasting alleles at a locus in the parent population are p and q, respectively, then $p + q = 1$, always; and genotype frequency in the progeny is $p^2 + 2qp + q^2 = 1$ or $p^2 + 2qp + q^2 = 1$.

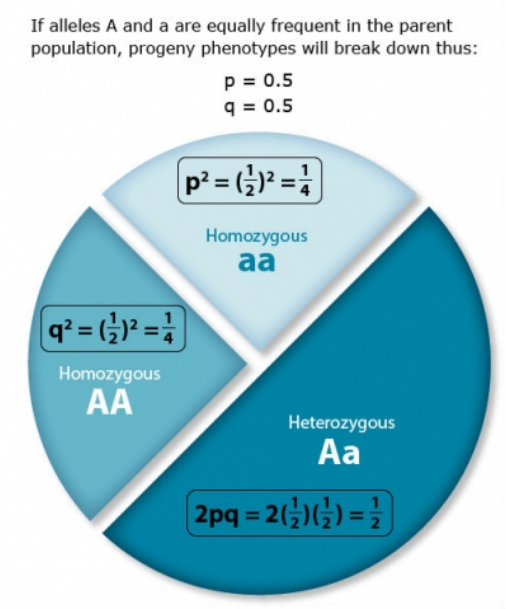


Fig. 9 Genotype frequencies in a hypothetical population in Hardy-Weinberg equilibrium.

Study Question 2 Explanation

The correct frequency of aa genotype following selection and random mating is 0.17. Selection for the A_ phenotype (or against the aa phenotype), shifts the allele and genotype frequencies. Here's how the answer is determined:

- Initial population is 0.09 AA + 0.42 Aa + 0.49 aa
- Selection removes aa genotypes, so the unselected portion of the population is 0.09 AA + 0.42 Aa and the remaining individuals are all A_.
- Thus, setting p equal to the frequency of the **A** allele, and q equal to frequency of the **a** allele, the resulting allelic frequencies are now

$$\mathrm{p} = \frac{\text{frequency of A in the AA genotype} + \text{frequency of A in the Aa genotype}}{\text{total allele frequencies of A and a}}$$

$$p = \frac{0.09 \times 2 + 0.42 \times 1}{0.09 \times 2 + 0.42 \times 2}$$

$$q = 1 - p = 0.41 \; q = 1 - p = 0.41$$

- So, the frequency of the **A** allele is 0.59 and the frequency of the **a** allele is 0.41.
- Now, we can calculate the frequency of the aa genotype in the population after one generation of selection and subsequent random mating.

p^2(AA) + 2pq(Aa) + q^2(aa) = 1
$(0.59)^2 + 2 \cdot 0.59 \cdot 0.41 + (0.41)^2 = 1$
0.35 AA + 0.48 Aa + 0.17 aa = 1

Thus, the correct frequency of the aa genotype is 0.17.

Factors Affecting Equilibrium

Several factors may disturb the genetic equilibrium of a population.

- Mutation of an allele at the locus of interest.
- Natural or human selection may favor one allele over the other.

- Migration of alleles into or out of the population (for example, via an introduction of a different allele from another population, or loss of an allele through selection).

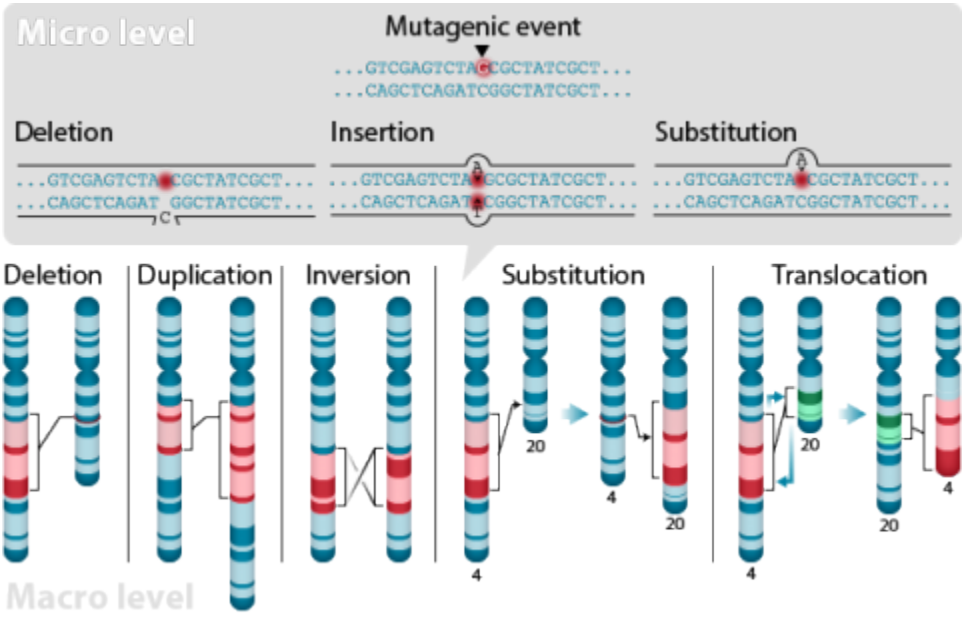


Fig. 10 Types of mutational events. Illustration by NIH/NHGRI, 2003.

Generally, a population not in genetic equilibrium, but retaining two contrasting alleles at a single, independently-segregating (non-linked) locus, will be restored to equilibrium at that locus after just one generation of random mating.

Random-Mating Interference

What is the significance of the Hardy-Weinberg Law to plant breeders? The random-mating assumption is often violated in breeding populations because breeding populations are smaller than natural plant populations. Thus, a mating design that minimizes gamete (allele) sampling errors is an important consideration. The breeder must be aware of several factors:

- **Self-pollinated population** — allele frequency will remain in equilibrium (assuming a sufficiently large population, no selection, or other factors that disturb equilibrium). However, with each successive generation of self-pollination, the genotype frequency of homozygous loci will increase and the frequency of heterozygous loci will decrease. Ultimately, the heterozygous genotype will be eliminated from the population with continued selfing.

- **Cross-pollinated population** — sampling errors occur if plants in the population differ in their vigor, time of flowering, or mate more frequently with plants in close proximity.
- **Selection for or against a particular allele will alter the allele and genotype frequencies of the population.** Selection against a dominant allele (i.e., selection for homozygous recessive) will remove the dominant allele from the population in a single generation. Selection against a recessive allele will require more than a few generations to remove the recessive allele from the population because the homozygous dominant and heterozygous genotypes have indistinguishable phenotypes.

Scenarios

In addition to being able to estimate allele and genotype frequencies, the breeder also needs to understand the gene action affecting the character of interest.

The breeding of cross-pollinated crops differs from self-pollinated species because of differences in the structures of their gene pools and opportunity for genetic recombination.

Table 1 Natural genetic structure of self- vs. cross-pollinated species.

Reproductive mode	Individuals	Population
Self-pollinated	Homozygous	Homogeneous or heterogeneous
Cross-pollinated	Heterozygous	Heterogeneous

Homozygosity and Heterozygosity

For a given locus, an individual with a genotype of either **AA** or **aa** is homozygous for that gene and is known as a homozygote; the status of the gene is referred to as homozygosity. An individual with the genotype **Aa** is heterozygous for that gene and is called a heterozygote; the status is known as heterozygosity. In the case of polyploid individuals, those with the genotypes **AAAA** (tetraploid) or **aaa** (triploid) would be examples of homozygotes and those with genotypes of **AAaa** (tetraploid) or **AAaaaa** (hexaploid) would be examples of heterozygotes.

The terms homozygous and heterozygous are used to describe the status of single genes or all gene loci within an individual, not within a population. There may be many different alleles of a

gene present in a population of individuals, but for each diploid individual, there are only two alleles per gene. For each individual, there is one allele from each parent and each allele per gene is present at corresponding loci on homologous chromosomes.

With regard to populations, a homogeneous population would be one in which all individuals in the population would have the same genotype and possess the same alleles for one or more genes. In contrast, a heterogeneous population would be characterized by differing alleles at one or more loci.

Note that a cross between two homozygous parents produces progeny that are homogeneous because all of the individual offspring are genetically identical. However, the offspring would be heterozygous for all loci for which different alleles occurred in the two parents.

Maize, the crop found in the first challenge, **Scenario 1—Fate of a Transgene**, is monoecious and is cross-pollinated.

Wheat, the crop found in the second challenge, **Scenario 2—Fixation of an Allele**, has bisexual flowers and is normally a self-pollinated crop.

Mating Systems for Crop Species

Table 6

Flower Type	Self-Compatability or Dioecy/ Monoecy	Crop Examples
Normally Cross-Pollinated		
Bisexual Flowers	Self-Compatible	Sugarcane, Olive, Amaranth, Avocado, Onion, Carrot, Agave, Sunflower, Kiwi, Pearl millet, Reed canarygrass, Sweet potato
Bisexual Flowers	Self-Incompatible	Radish, Kale, Cabbage, Black mustard, Pineapple, Red Clover, White Clover, Apple, Pear, Cacao, Rye, Alfalfa, Birdsfoot trefoil, Sweet Potato, Buckwheat
Unisexual Flowers	Dioecious	Papaya, Fig, Hops, Hemp, Grape
Unisexual Flowers	Monecious	Mango, Cucumber, Squash, Watermelon, Yam, Rubber, Cassava, Castor bean, Maize, Banana, Coconut, Oil palm
Normally Self-Pollinated		
Bisexual Flowers	Self-Compatible	Barley, Oats, Rice, Triticale, Wheat, Lettuce, Cowpea, Dry bean, Lentil, Chickpea, Peanut, Pea, Soybean, Sesame, Tomato, Tobacco, Coffee, Eggplant, Safflower, Flax, Peach
Predominantly Self-Pollinated, but also Cross-Pollinated to fairly high extent		
Bisexual Flowers	Self-Compatible	Cotton, Sorghum, Rapeseed, Brown mustard

Let's examine the genetic structure of populations of self- and cross-pollinated species.

Scenario 1—Fate of a Transgene

Imagine a community of small farms in a valley located in the highlands of Central America. The farmers of this community produce grain from an open-pollinated maize variety that is adapted to their preferred cultural practices. They also select partial ears from about 5% of their better performing plants to be used for seed in their next growing season. One day a truck filled with seed of a transgenic hybrid overturns on the highway while passing through the valley. 99.999% of the seed is recovered, but about 500 kernels remain in a farmer's 10-acre

field adjacent to the highway. The transgenic seeds germinate and grow to maturity alongside the planted open-pollinated variety. You are asked to determine the fate of an insect-resistant transgene in this valley.

Scenario 2—Fixation of an Allele

Imagine a naturally occurring allele at a locus that regulates the structure of carbohydrates in the wheat kernel; with the allele, the carbohydrates in the kernel have low glycemic indices. For the last 100 years, hard-red winter wheat varieties have not been selected for low glycemic indices, but with the emergence of a Type II diabetes epidemic, there is a demand for low glycemic carbohydrates in hard-red winter wheat varieties. How will you develop a breeding population in which this allele is fixed, that is the frequency of this allele = 1.0?

Genetics of Cross-Pollinated Species

Because cross-pollinated species have evolved to outcross, individuals tend to be heterozygous at many loci and they usually perform best when that heterozygosity is maintained. This is a characteristic referred to as heterosis or hybrid vigor. When repeated self-pollination occurs in cross-pollinated species, homozygosity increases and plant vigor is reduced, a phenomenon called inbreeding depression. Heterosis and inbreeding depression will be further discussed in Lesson 6.

Several morphological and physiological features of cross-pollinated species promote cross-pollination. Let's briefly review these.

- **Monoecy** — pistillate and staminate flowers occur on different sections of the same plant.
- **Dioecy** — pistillate and staminate flowers occur on different plants.
- **Protandryorprotogyny** — pistillate and staminate flowers mature at different times.
- **Self-incompatibility** — pollen from the same plant cannot effect fertilization or seed set.
- **Male or female sterility** — pollen or ovule does not function normally.

Genetics of Self-Pollinated Species

Self-pollinated species rarely hybridize naturally. Although cross-pollinating may occasionally occur, ovules of a self-pollinated plant are normally fertilized by pollen produced on that

same plant. The result of repeated generations of selfing is that homozygosity is increased or maintained.

- Homozygous loci will remain homozygous.
- Heterozygous loci will segregate such that the frequency of homozygotes will increase at the expense of the frequency of heterozygotes with each generation of selfing.

Frequency of Homozygotes

With continued self-pollination, the heterozygotes will segregate, decreasing the proportion of heterozygotes in the population by half each generation. Notice that the homozygotes can only produce homozygotes.

Table 5 Change in percent heterozygosity in each successive generation.

Generation	Heterozygosity (%)
F_1	100.0
F_2	50.0
F_3	25.0
F_4	12.5
F_5	6.25
F_6	3.12

For each successive generation of offspring resulting from one F1 individual, by the F8 generation, the population is essentially homozygous. When no further segregation for the trait occurs, all progeny derived from that F1 will "breed true" because they are homozygous for the trait. The proportion of plants that are expected to be heterozygous at any gene when starting with a heterozygous F_1 and selfing can be determined by using the formula $(½)^n$, where n = the number of segregating generations, e.g., in F_2 n = 1 and in F_5 n = 4.

Proportion of homozygous plants in any generation is then given by $1 - (½)^n$ which when algebraically converted is equal to: $\frac{2^n - 1}{2^n}$

How does a locus become heterozygous? A contrasting allele can be acquired when a plant out-crosses or when a mutation occurs. Each successive self-pollination thereafter will reduce heterozygosity by half. Breeders rely on the natural tendency of self-pollinated crops to become homozygous to obtain lines that exhibit uniformity in characters that affect appearance and performance.

Notice how rapidly populations lose heterozygosity with selfing. For self-pollinated crops, one of the breeder's objectives is usually to develop pure lines. Since pure lines are homozygous, their rapid loss of heterozygosity speeds cultivar development. Some background heterozygosity may remain in a pure line, but the line is sufficiently homozygous to provide the uniformity in characters required for reliable and predictable appearance and performance.

Allelic Effects

The tendency of a species to self-pollinate or outcross influences allelic and genotypic frequencies in the population. In a self-pollinated homozygous population, the effect of a gene (allele) is determined by the gene's effect in combination with itself and with alleles at other loci. What determines the effect of a gene in a cross-pollinated population?

Effect or fate of an allele in a cross-pollinated population is determined by its effect

- in combination with other alleles at the same locus
  - additive effects
  - dominance effects
  - overdominance effects
- in combination with alleles at other independent loci (epistatic effects)
- in combination with alleles at closely linked loci

One difference between a self-pollinated and a cross-pollinated population is that in the cross-pollinated population there is constant inter-crossing. Thus, recombination and rearrangement of alleles and expression of dominance and epistatic effects occur.

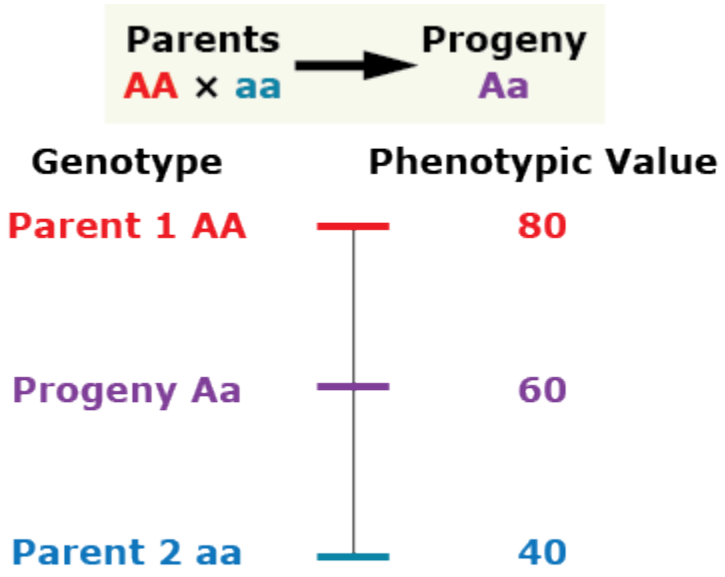
Review **gene action** or **gene interactions**, such as epistasis in the next screens.

Gene Action

There are several general types of gene action. The type of gene action and the alleles present

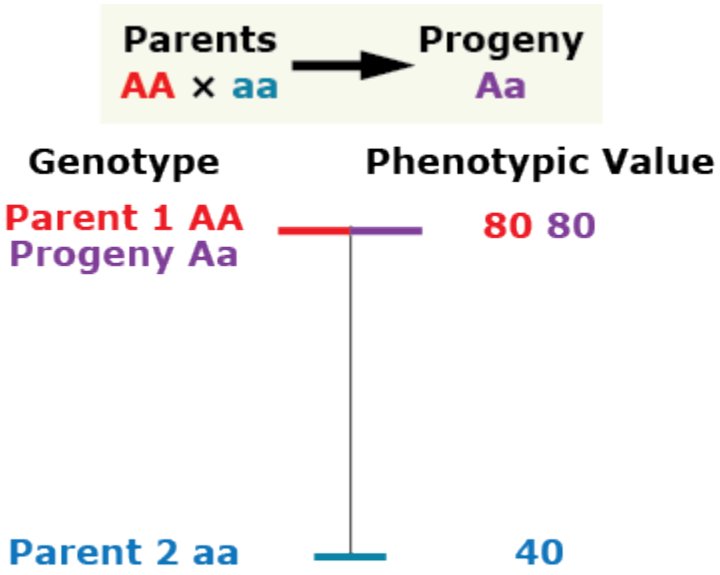
for a given gene affect the phenotype. Let's consider the gene action as indicated by the phenotype of a diploid individual heterozygous at the given single locus compared to the phenotype of its parents.

Additive gene action (no dominance)



The progeny's phenotypic value is at the midpoint between both parents. In this example, each A allele adds 20 units.

Complete Dominance



The phenotype of the heterozygous progeny equals the phenotype of the homozygous dominant parent.

Partial (incomplete) dominance

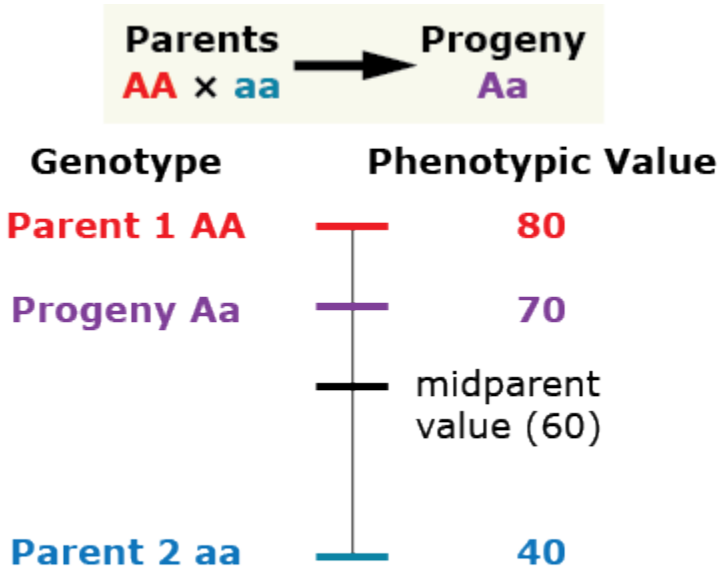


Fig. 11. The heterozygous progeny has a phenotypic value greater than that of the midparent value, but less than that of the homozygous dominant parent.

Over-Dominance

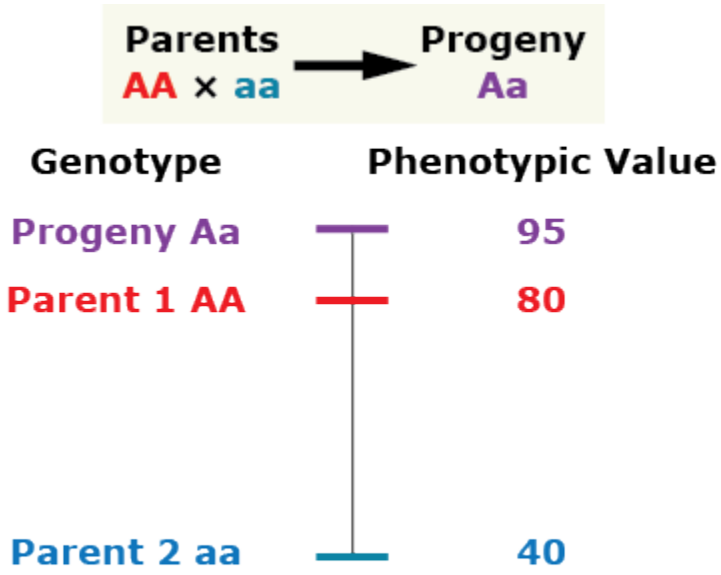


Fig. 12. The phenotype of the heterozygous progeny is greater than either parent.

Gene Interactions

When multiple genes control a particular trait or set of traits, gene interactions can occur. Generally, such interactions are detected when genetic ratios deviate from common phenotypic or genotypic proportions.

- **Pleiotropy** — Genes that affect the expression of more than one character.
- **Epistasis** — Genes at different loci interact, affecting the same phenotypic trait.

Epistasis occurs whenever two or more loci **interact to create new phenotypes**. Epistasis also occurs whenever an allele at one locus either **masks** the effects of alleles at one or more other loci or if an allele at one locus **modifies** the effects of alleles at one or more other loci. There are numerous types of epistatic interactions.

Epistasis is expressed at the **phenotypic** level. It is important to note that genes that are involved in an epistatic interaction may still exhibit independent assortment at the **genotypic** level. In the case of two completely dominant, non-interacting (i.e., no linkage) genes, all of the

deviations observed in results involving epistatic interactions are modifications of the expected 9:3:3:1 ratio.

Proof

The proof of Hardy-Weinberg Equilibrium (HWE) requires the following assumptions (Falconer and Mackay, 1996):

1. Allele frequency in the parents is equal to the allele frequency in the gametes
    a. Assumes normal gene segregation
    b. Assumes equal fertility of parents
2. Allele frequency in gametes is equal to the allele frequency in gametes forming zygotes
    a. Assumes equal fertilizing capacity of gametes
    b. Assumes large population
3. Allele frequency in gametes forming zygotes is equal to allele frequencies in zygotes
4. Genotype frequency in zygotes is equal to genotype frequency in progeny
    a. Assumes random mating
    b. Assumes equal gene frequencies in male and female parents
5. Genotype frequencies in progeny do not alter gene (allele) frequencies in progeny.
    a. Assumes equal viability

For a two allele locus in a population in HWE: $P_{AA} = p^2; P_{Aa} = 2_{pq}; P_{aa} = q^2$

Proof

HWE at a given genetic locus is achieved in one generation of random mating. Genotype frequencies in the progeny depend only on the gene (allele) frequencies in the parents and not on the genotype frequencies of the parents.

If a population is in HWE, relationships between frequencies of alleles and genotypes may be derived as depicted in figure 13.

As shown in figure 13, in HWE:

- frequency of heterozygotes does not exceed 0.5
- heterozygotes are most frequent genotype when p or q are between 0.33 and 0.66
- very low allele frequency should result in very low frequency of homozygotes for that allele
- if there are only two alleles at a locus in the population, p+q=1.

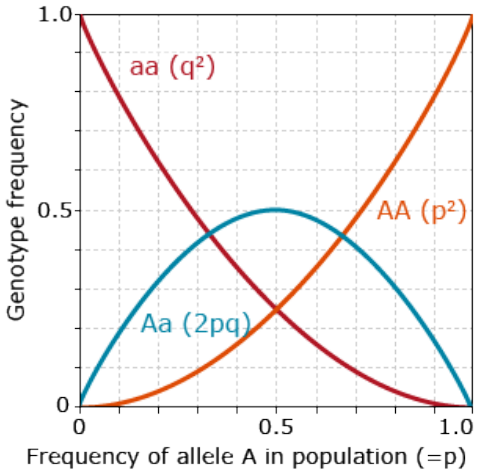


Fig. 13 Relationship between genotype frequencies and gene frequency for two alleles in a population in Hardy-Weinberg equilibrium.

A **chi-square test** is typically used to determine whether or not a population varies significantly from Hardy-Weinberg expectations. The Hardy-Weinberg formula is useful in describing situations where mating is completely randomized. But more commonly, mating is not at random and populations are subjected to other forces, such as mutation, migration, genetic drift, and selection. Linkage can also have a significant effect on gene frequencies.

Forces Affecting Population Structures

Descriptions

Non-Random Mating

Two methods of non-random mating that are important in plant breeding are **assortative mating** and **disassortative mating**.

Assortative mating occurs when similar phenotypes mate more frequently than they would by chance. One example would be the tendency to mate early x early-maturing plants and late x late maturing plants. The effect of assortative mating is to increase the frequency of homozygotes and decrease the frequency of heterozygotes in a population relative to what would be expected in a randomly mating population. Assortative mating effectively divides the population into two or more groups where matings are more frequent within groups than between groups.

Disassortative mating occurs when unlike or dissimilar phenotypes mate more frequently than would be expected under random mating. Its consequences are in general opposite those of assortative mating in that disassortative mating leads to an excess of heterozygotes and a deficiency of homozygotes relative to random mating. Disassortative mating can also lead to the maintenance of rare alleles in a population. For example, in self-incompatible species, an individual will only mate with another individual that differs in the self-incompatibility loci. This is a type of disassortative mating, resulting in a great alleleic diversity in the self-incompatibility loci. It is an effective mechanism to maintain heterozygosity and prevent inbreeding.

Forces Affecting Allele Frequency

Factor Categories

The factors affecting changes in allele frequency can be divided into two categories: **systematic processes**, which are predictable in both magnitude and direction, and **dispersive processes**, which are predictable in magnitude but not direction. The three systematic processes are migration, mutation, and selection. Dispersive processes are a result of sampling in small populations.

Table 2

Systematic Processes	Dispersive Processes
Migration	Small Population Size
Mutation	
Selection	

Migration

Clearly, the first challenge described in the introduction represents a case of migration. A new set of genes in a developed transgenic hybrid have been introduced into an open pollinated variety of maize. When discussing population genetics, migration is also sometimes referred to as **gene flow**, a concept that is often used interchangeably with migration by population geneticists. However, the term migration means the movement of *individuals* between populations, whereas gene flow is the movement of *genes* between populations. New genes

would be established in the population if the immigrant successfully reproduces in its new environment, but if it doesn't reproduce migration would still have occurred while gene flow would not.

Assume a population has a frequency of m new immigrants each generation, with 1– m being the frequency of natives. Let q_m be the frequency of a gene in the immigrant population and q_0 the frequency of that gene in the native population. Then the frequency in the mixed population will be:

$$q_1 = mq_m + (1 - m)q_0$$

$$q_1 = m(q_m - q_0) + q_0$$

The change in gene frequency brought about by migration is the difference between the allele frequency before and after migration

$$\Delta q = q_1 - q_0$$

$$\Delta q = m(q_m - q_0)$$

Thus the change in gene frequency from migration is dependent on the rate of migration and the difference in allele frequency between the native and immigrant population. Migration or gene flow can introduce new alleles into a population at a rate and at more loci than expected from mutation. It can also alter allele frequencies if the populations involved have the same alleles but not in the same proportions. Thus the effect of migration on changes in allele frequency depends on differences in allele frequencies (migrants vs. residents) and the proportion of migrants in the population.

Mutation

Mutations are the source of all genetic variation. Loci with only one allelic variant in a breeding population have no effect on phenotypic variability. While all allelic variants originated from a mutational event, we tend to group mutational events in two classes: rare mutations and recurrent mutations where the mutation occurs repeatedly.

Rare Mutations

By definition, a rare mutation only occurs very infrequently in a population. Therefore, the

mutant allele is carried only in a heterozygous condition and since mutations are usually recessive, will not have an observable phenotype. Rare mutations will usually be lost, although theory indicates rare mutations can increase in frequency if they have a selective advantage.

Fate of a Single Mutation

Consider a population of only **AA** individuals. Suppose that one **A** allele in the population mutates to **a**. Then there would only be one **Aa** individual in a population of **AA** individuals. So the **Aa** individual must mate with a **AA** individual.

AA x Aa → 1AA:1Aa

From Li (1976; pp 388), this mating has the following outcomes:

1. **No offspring are produced** in which case the mutation is lost.
2. **One offspring is produced:** the probability of that offspring being **AA** is 1/2 so the probability of losing the mutation is 1/2.
3. **Two offspring are produced:** the probability of them both being **AA** is 1/4 so the probability of losing the mutation is 1/4.

If k is the number of offspring from the above mating then the probability of losing the mutation among the first generation of progeny is $(1/2)^k$.

The probability of losing the gene in the second generation can be calculated by making the following assumptions:

- Number of offspring per mating is distributed as a Poisson process (which means that they follow a stochastic distribution in which events occur continuously and independently of one another).
- With the average number of offspring per mating = 2.
- New mutations are selectively neutral.

With these assumptions, the probabilities of extinction are:

Table 3 Probability of extinction in different generations.

Generation	Probability of Loss
1	0.37
7	0.79
15	0.89
31	0.94
63	0.97
127	0.98

Recurrent Mutations

Let the mutation frequencies be:

Mutation rate: $A \overset{u}{\rightarrow} a$

Frequency: $p_0 \overset{\leftarrow}{v} q_0$

Then the change in gene frequency in one generation is:

$$\Delta q_0 = up_0 - vq_0$$

at equilibrium

$$p_0 u = q_0 v$$

$$q_0 = \frac{u}{v+u}$$

Conclusions:

- Mutations alone produce very slow changes in allele frequency
- Since reverse mutations are generally rare, the general absence of mutations in a population is due to selection

Selection

Selection is one of the primary forces that will alter allele frequencies in populations. **Selection** is essentially the differential reproduction of genotypes. In population genetics, this concept is referred to as **fitness** and is measured by the reproductive contribution of an individual (or genotype) to the next generation. Individuals that have more progeny are more fit than those who have less progeny because they contribute more of their genes to the population.

The change in allele frequency following selection is more complicated than for mutation and migration, because selection is based on phenotype. Thus, calculating the change in allele frequency from selection requires knowledge of genotypes and the degree of dominance with respect to fitness. Selection affects only the gene loci that affect the phenotype under selection—rather than all loci in the entire genome—but it also would affect any genes that are linked to the genes under selection.

Effects of Selection

Change in allele frequency

The strength of selection is expressed as a **coefficient of selection**, **s**, which is the proportionate reduction in gametic output of a genotype compared to a standard genotype, usually the most favored. **Fitness** (relative fitness) is the proportionate contribution of offspring to the next generation.

Partial selection against a completely recessive allele

To see how the change in allele frequency following selection is calculated consider the case of selection against a recessive allele:

Table 4

	Genotypes			
	AA	**Aa**	**aa**	**Total**
Initial Frequencies	p^2	2pq	q^2	1
Coefficient of Selection	0	0	s	
Fitness	1	1	1-s	
Gametic Contribution	p^2	2pq	$q^2(1-s)$	$1 - sq^2$

Frequency Equations

The frequency of allele **a** after selection is:

$$q_1 = \frac{q - sq^2}{1 - sq^2}$$

$$q_1 = \frac{q - sq^2}{1 - sq^2}$$

The change in allele frequency is then:

$$\Delta q = q_1 - q$$

$$\Delta q = \frac{q - sq^2}{1 - sq^2} - q$$

In general, you can show that the number of generations, t, required to reduce a recessive from a frequency of q_0 to a frequency of q_t, assuming complete elimination of the recessive (s = 1) is:

$$t = \frac{1}{q_t} - \frac{1}{q_0}$$

Discussion

Review the two challenges at the beginning of the lesson and then answer these questions:

- **Allele Frequency**—For Scenario 1, **calculate the frequency** of the insect-resistant transgene in the Central American maize farmer's 10-acre field assuming that it is a) hemizygous and b) homozygous in the spilled hybrid seed. Remember that hemizygous means that the individual has only one single homologous chromosome, and therefore is neither homozygous nor heterozygous; in contrast homozygous means that there are two homologues.
- **Allele Frequency**—For Scenario 2, **calculate the frequency** of the allele responsible for

low glycemic carbohydrates in the wheat breeding population, assuming the allele is not present in any wheat variety except one.

- **Mutation**—For Scenario 2, assume the mutation that produced the low glycemic allele was selectively neutral in the hard red-winter wheat breeding population. **Why was that allele lost from all varieties that were developed over the last 100 years?**
- **Selection**—In Scenario 1 a transgene (and likely other genes) is introduced into an open-pollinated variety in one farmer's field. **Determine Δq for the transgenic allele** assuming that the allele is homozygous in the hybrid seed, the insect-resistant allele is completely dominant and the selective advantage of the allele is a) two to one (2:1) when the insect is present and b) one to one (1:1) when it is absent.

Scenario 1: Fate of a Transgene

Imagine a community of small farms in a valley located in the highlands of Central America. The farmers of this community produce grain from an open-pollinated maize variety that is adapted to their preferred cultural practices. They also select partial ears from about 5% of their better performing plants to be used for seed in their next growing season. One day a truck filled with seed of a transgenic hybrid overturns on the highway while passing through the valley. 99.999% of the seed is recovered, but about 500 kernels remain in a farmer's 10-acre field adjacent to the highway. The transgenic seeds germinate and grow to maturity alongside the planted open pollinated variety. You are asked to determine the fate of an insect-resistant transgene in this valley.

Scenario 2: Fixation of an Allele

Imagine a naturally occurring allele at a locus that regulates the structure of carbohydrates in the wheat kernel; with the allele the carbohydrates in the kernel have low glycemic indices. For the last 100 years hard-red winter wheat varieties have not been selected for low glycemic indices, but with the emergence of a Type II diabetes epidemic, there is a demand for low glycemic carbohydrates in hard-red winter wheat varieties. How will you develop a breeding population in which this allele is fixed, that is the frequency of this allele = 1.0?

Small Population Size

Unlike the three systematic forces that are predictable in both amount and direction, changes due to small population size are predictable only in amount and are random in direction.

The effects of small population size can be understood from two different perspectives. It can be considered a sampling process and it can be considered from the point of view of inbreeding. The inbreeding perspective is more interesting, but looking at it from a sampling perspective lets us understand how the process works.

A particular sub-population is a random sample of N individuals or 2N gametes (for a diploid) from the base population. Therefore, the expected gene frequency of a particular allele in the sub-populations is q_0 and the variance of q is $\sigma_q^2 = \frac{p_0 q_0}{2N}$

Since q_0 is a constant, the variance of the change in allele frequency ($q_1 - q_0$) is also: $\sigma_{\Delta q}^2 = \frac{p_0 q_0}{2N}$

Examples

Example 1: Let q = 0.5 and N = 50, then $\sigma_q^2 = \frac{(0.5)(0.5)}{100} = 0.0025$

Example 2: Let q = 0.5 and N = 4, then $\sigma_q^2 = \frac{(0.5)(0.5)}{8} = 0.03125$

Consequences of small population size

1. Random genetic drift: random changes in allele frequency within a subpopulation
2. Differentiation between subpopulations
3. Uniformity within subpopulations
4. Increased homozygosity

Random Genetic Drift

Small Population Size

Random genetic drift refers to allelic frequencies that change through time (generations) due to errors and other random factors (i.e., not selection or mutation). When sample sizes are small, all genotypes may not be produced and then mate at expected frequency. The **effective population** size ($\mathbf{N_e}$) of a population is a term used to describe the number of parents that actually contribute gametes to the next generation; not all individuals may contribute equally, thus resulting in genetic drift. Small populations are susceptible to **genetic bottlenecks**, which are sudden decreases in breeding population due to deaths, migration, or other factors. Small populations can be subject to so-called **founder effects**, which occur when a breeding population is small when initially founded, then increases in size but the gene pool is largely determined by the genes present in the original founders.

Rate of Change

The rate of change due to random genetic drift depends on population size and allele frequency. As illustrated in the figure below, the more frequent the allele, the higher chances of being fixed and the smaller the population, the faster it will either move towards fixation or loss. In the absence of other forces:

- genetic drift leads to loss or fixation of alleles
- frequency of rare alleles would be expected to go to zero
- lower frequency of heterozygotes in later generations
- less genetic variation within subpopulations
- more genetic variation among subpopulations

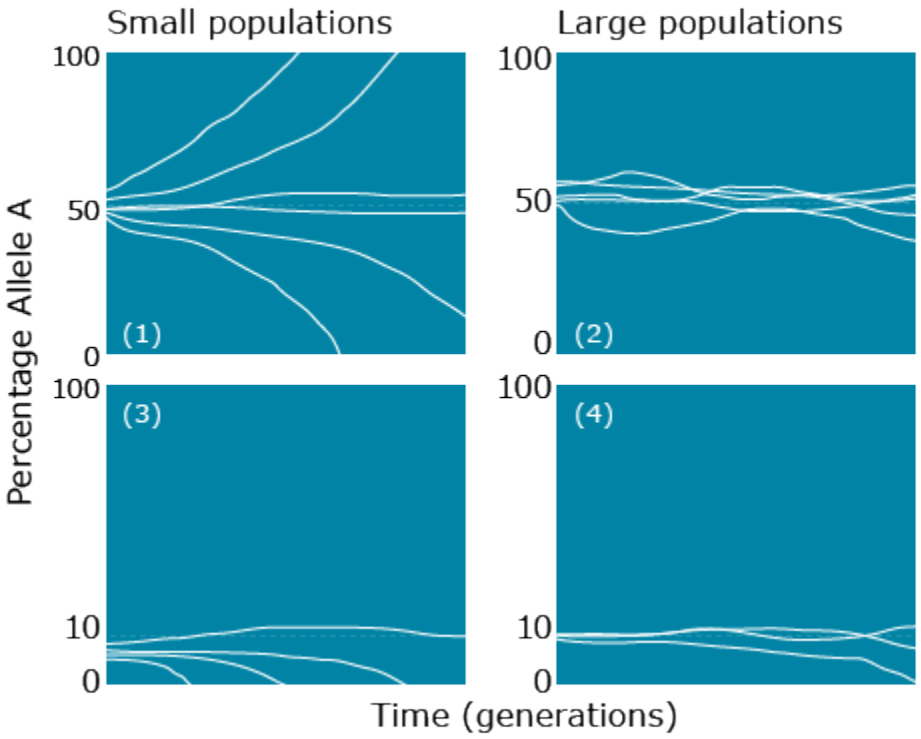


Fig. 15 Effect of population size and gene frequency on rate of fixation due to drift. Each line represents a different population. In small populations, allele frequency for A showed greater differences among populations (1) whereas in larger populations, allele frequencies are similar over generations (2). When allele frequency is very low, the allele is more likely to be lost in small populations (3) than in larger populations (4).

Inbreeding and Small Populations

Inbreeding and Small Populations Inbreeding is the mating together of individuals that are related by ancestry. The degree of relationship among individuals in a population is determined by the size of the population. This can be seen by examining the number of ancestors that a single individual has:

Just 50 generations ago note that a single individual would have more ancestors than the number of people that have existed or could exist on earth.

Therefore, in small populations individuals are necessarily related to one another. Pairs mating at random in a small population are more closely related than pairs mating together in a large population. Small population size has the effect of forcing relatives to mate even under random mating, thus with small population sizes inbreeding is inevitable.

Generation	Ancestors
0	1
1	2
2	4
3	8
4	16
5	32
6	64
10	1,024
50	1,125,899,906,842,620
100	1,267,650,600,228,230,000,000,000,000,000
t	2^t

Identical Types

In finite populations there are two sorts of homozygotes: Those that arose as a consequence of the replication of a single ancestral gene — these genes are said to be identical **by descent** (Bernardo, 1996). If the two genes have the same function, but did not arise from replication of a single ancestral gene, they are said to be alike **in state**. It is the production of homozygotes that are identical by descent that gives rise to inbreeding in a small population.

Summary of Factors

Hardy and Weinberg discovered mathematically that genotype frequencies will reach an equilibrium in one generation of random mating in the absence of any other evolutionary force. If the conditions of equilibrium are met, the frequencies of different genotypes in the progeny will depend only upon the allele frequencies of the previous generation. If allele frequencies do not accurately predict genotype frequencies, then plants are mating in a non-random way or another evolutionary force is operating.

	Effect of level of variation		
	Within subpopulations	**Among subpopulations**	**Affect all loci equally?**
Mutation	Increase	Increase	No
Migration (Gene flow)	Increase	Decrease	Yes
Random Genetic Drift	Decrease	Increase	Yes
Selection	Increase or decrease	Increase or decrease	Yes

Within subpopulations the degree of genetic variation can be assessed by heterozygosity, while variation among subpopulations is measured by population differentiation. Mutation is the ultimate source of all genetic variation and it tends to increase variation both within and among subpopulations. But because most mutations are rare, the effect of mutation is slow relative to the change the other forces can effect. Migration or gene flow and random genetic drift are opposite in their effects: migration tends to increase variation within subpopulations but decrease it among subpopulations, and random drift does the opposite. In contrast, the effects of selection vary both within and between populations. For example, variation can decrease if one homozygote is favored, or may increase or be maintained if heterozygosity is advantageous. Selection acts on the phenotype so it will affect only those genes that control the trait under selection, as well as genes linked to those loci.

Schematic Overview

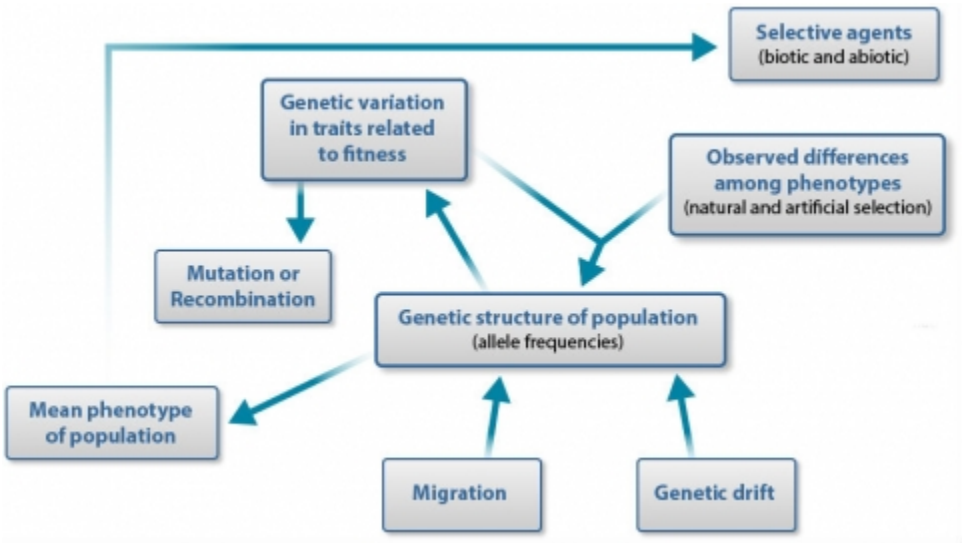


Fig. 16 Schematic overview of key concepts in population genetics. Source: Conner and Hartl, 2004.

References

Bernardo, R., A. Murigneux, and Z. Karaman. 1996. Marker-based estimates of identity by descent and alikeness in state among maize inbreds. Theoretical and Applied Genetics 93: 262-267.

Conner, J. K., and D.L. Hartl. 2004. A Primer of Ecological Genetics. Sinauer Associates, Sunderland, MA.

Falconer, D.S. and T.F.C. Mackay. 1996. Introduction to Quantitative Genetics. 4th edition. Longman Pub. Group, Essex, England.

Hancock, J.F. 2004. Plant Evolution and the Origin of Crop Species. 2nd edition. CABI Publishing, Cambridge, MA.

National Institutes of Health. National Human Genome Research Institute. "Talking Glossary of Genetic Terms." http://www.genome.gov/glossary/

Pierce, B. A. 2008. Genetics: A Conceptual Approach. 3rd edition. W.H. Freeman, New York.

How to cite this chapter: Beavis, W., L. Merrick, K. Meade, A. Campbell, D. Muenchrath, and S. Fei. 2023. Population Genetics. In W. P. Suza, & K. R. Lamkey (Eds.), *Crop Genetics.* Iowa State University Digital Press. DOI: 10.31274/isudp.2023.130

Inbreeding and Heterosis

Laura Merrick; Jode Edwards; Thomas Lübberstedt; Arden Campbell; Deborah Muenchrath; Shui-Zhang Fei; and William Beavis

Introduction

This module focuses on **inbreeding**, a type of mating of individuals that is often of particular significance to plant breeders. Inbreeding is defined as the mating of individuals that are related by ancestry. Self-pollination (mating of an individual to itself) represents the most extreme form of inbreeding. Inbreeding leads to an increase in homozygosity at the expense of heterozygosity. A key feature of inbreeding is that as homozygosity increases in a population undergoing inbreeding in the absence of selection, the genotype frequency changes while the allele frequency stays unchanged. Inbreeding may occur unintentionally as a result of selection or maintenance of small populations. Inbreeding is also deliberately practiced as a method to create genetic uniformity in populations of interest for genetic or breeding research, for retaining genotypes of inbred cultivars of self-pollinated species through many years of production, or for reliable production of **inbred lines** to be used in the development of commercial **hybrid cultivars**.

Inbreeding Depression

A phenomenon known as **inbreeding depression**—the reduced survival and fertility of offspring of related individuals—occurs in both plants and animals, showing that variation for heritable fitness traits occurs within populations. The occurrence of inbreeding depression varies across species. Charles Darwin, British naturalist famous for his theories of evolution and natural selection was the first person to make a distinction between plants that are outbreeders (species with reproductive mechanisms promoting cross-pollination, who typically exhibit inbreeding depression and tend to be intolerant of inbreeding, e.g., maize and alfalfa) and inbreeders (species with reproductive mechanisms promoting self-pollination in which inbreeding depression is minimal, who tolerate many generations of inbreeding, e.g., wheat and oat). Darwin noticed that:

- outcrossing is more common in nature than self-fertilization
- there are complex reproductive systems that promote outcrossing in plants
- many plant species have evolved systems that prevent self-fertilization.

This module also describes hybrid vigor or **heterosis**, which is a phenomenon that is functionally the opposite of inbreeding depression. Heterosis is defined as the increased vigor of F_1 progeny resulting from the mating of inbred parents. Generally, the performance of the F_1 hybrid exceeds the performance of its inbred parents for various traits. The expression of hybrid vigor might affect phenotypes under single gene or polygenic control, e.g., size, growth rate, fertility, and yield. The first-generation offspring of crosses from mating between different pure-line inbreds generally show in a greater amount of desired traits of both parents, but hybrid vigor decreases in F_2 and subsequent selfing generations due to inbreeding. The exploitation of hybrid vigor or heterosis is a key feature in the success of hybrid cultivars.

Cross-Pollination

Several mechanisms promote cross-pollination.

- Emergence or maturity of the staminate and pistillate flowers is asynchronous.
    - **Protandry** – anthesis occurs before stigma are receptive.
    - **Protogyny** – pistillate flower matures before the staminate flower.
- Flowers are monoecious or dioecious.
- Mechanical obstruction between the staminate and pistillate flowers in the same individual prevents self-pollination. Alfalfa flowers, for example, have a membrane over the stigma that precludes self-pollination. When a bee lands on the flower, the keel is tripped, rupturing the membrane and exposing the stigma to pollen carried by the bee from other plants it has visited, effecting cross-pollination.
- Gametes produced on the same plant or clone are unable to effect fertilization.
    - **Self-sterility** – gametes from same individual cannot successfully fuse to form a zygote. Sterility can be caused by lack of function of pollen (male gametes) or ovules (female gametes). **Male sterility**, either genetic or cytoplasmic, occurs because the pollen is not viable. Female sterility occurs when the ovule is defective or seed

development is inhibited.

- **Self-incompatibility** – self-pollination may occur, but fertilization and seed set fail.

Pollen is transported from the staminate flower to the pistillate flower by wind, insects, or animals. Occasionally pollen is transported to receptive stigma of the same individual and self-pollination may occur. For example, pollen from the tassel of a maize plant may land on and pollinate silks on the same plant, effecting self-pollination.

Sunflower is ordinarily cross-pollinated. Bees often carry pollen from one plant and deposit it on other plants.

Fig. 1 Beehives on the edge of a sunflower field. Photo by Iowa State University.

Fig. 2 Close-up of a sunflower head. Photo by Iowa State University.

Self-Pollination

Several floral mechanisms enforce self-pollination.

- Flowers do not open, preventing external pollen from reaching the stigma.
- **Anthesis** occurs before the flower opens.
- Stigma elongates through the staminal column (filaments and anthers) immediately after anthesis.
- Floral organs may obscure the stigma after the flower opens.

Although these mechanisms usually enforce self-pollination, a low frequency of cross-pollination may occur. The frequency of cross-pollination in normally self-pollinating species generally depends on the species and environmental conditions.

Soybean is an example of a species that is normally self-pollinated. Before the flower opens, the anthers burst and pollen grains fall out of the anthers on to the receptive stigma contained in the same flower-self-pollination occurs.

Objectives

- Understand the effects of inbreeding.
- Be able to assess the amount of inbreeding by consideration of the inbreeding coefficient.
- Compare and contrast systems of mating that promote inbreeding—self-pollination, half-sib mating, full-sib mating, and backcrossing.
- Learn why inbreeding depression occurs and know why it happens more commonly in species that are predominately cross-pollinated vs. those that are predominately self-pollinated.
- Understand the effects of heterosis and know the difference between the dominance vs. the overdominance hypotheses for explaining its occurrence.

Genetics of Inbreeding

Inbreeding as a Probability

Inbreeding is characterized by a departure from random mating and involves preferential mating between relatives. The most significant effect of inbreeding is that replicates of a single allele in a common ancestor can come together through the mating of relatives to produce homozygous progeny whose alleles are identical by descent. In contrast in a second type of homozygote, two alleles that are said to be alike in state are those that have the same function, but did not derive from replication of a single ancestral gene. These two types of homozygotes can be understood by analyzing pedigrees such as the ones shown in Fig.3. The two copies of the A_1 allele in the A_1A_1 offspring individual at the bottom of Fig. 3a descend from the same copy in a common ancestor; this type of genotype is termed autozygous. In contrast the two copies of the A_1 allele in the A_1A_1 offspring individual at the bottom in Fig. 3b are descended from two different copies in ancestors; this type of genotype is allozygous.

Inbreeding can be estimated directly by studying pedigrees or lineages of ancestry showing genetic relationships.

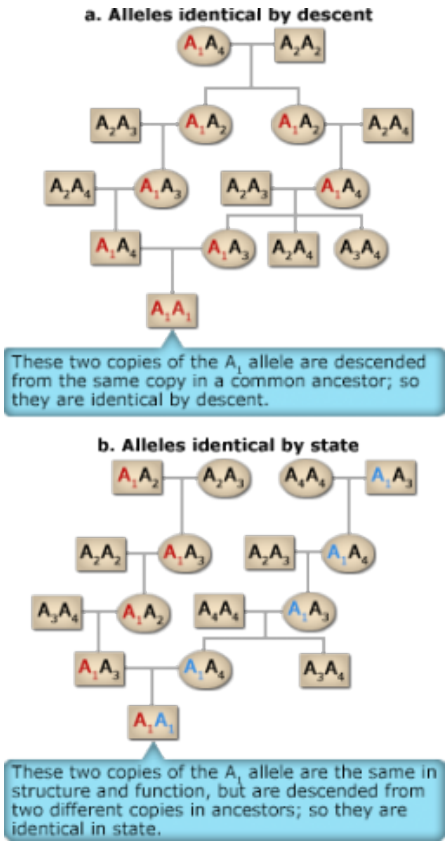


Fig. 3 Individuals homozygous by descent or by state.

Measurement of Inbreeding

The probability that two genes are identical by descent is called the coefficient of inbreeding (denoted as *F*) and will be the measure of relationship between mating pairs.

The coefficient of inbreeding (*F*) is defined as the probability that two alleles at the same locus are identical by descent. At the population level, *F* describes the average level of homozygosity. The coefficient of inbreeding is always expressed relative to a specified base population. The base population is defined to be non-inbred (*F*=0). The range of *F* is 0 to 1, with 0 indicating random mating and no inbreeding, while 1 means prolonged selfing.

Consider a base population consisting of N individuals each shedding equal numbers of gametes uniting at random (although see Fehr, 1987, p.113, with respect to the concept of effective population size, or N_e). Because the base is defined to have *F*=0, each individual in this population carries genes that are non-identical. The only way a homozygote that carries genes that are identical by descent can arise is by the mating of a male and female gamete from the same individual that carries a replication of the same gene. Because there are 2N gametes the probability that two mating gametes are identical by descent is 1 / 2N.

Effective Population Size

The concept of effective population size, denoted by the symbol N_e, was introduced by Sewall Wright as number of breeding individuals in an idealized (as small as possible, i.e., "redundant" individuals would be eliminated) population showing the same distribution of allele frequencies under random genetic drift or inbreeding as the population under consideration. The effective population size is usually smaller than the absolute population size (N). As noted in pages 112-114 in the Fehr textbook, N_e is a relative measure of the number of parents mating in a population >>> N_e (see Fehr, 1987, p. 112-114).

The general equation for N_e is $N_e = \frac{2N}{1 = F_p}$

where:

- **N** = number of individuals that are mating
- $\mathbf{F_p}$ = coefficient of inbreeding for parents
- $\mathbf{N_e}$ is a basic parameter in many models in population genetics, and will be explained in more detail in the Population and Quantitative Genetics for Breeding course.

Second Generation

In the second generation, there are two ways genes that are identical by descent can be joined: 1) by a new replication of the same ancestral gene; and 2) by the previous replication that occurred in generation 1. The probability of a new replication event is $\frac{1}{2N}$. The remaining

proportion of zygotes, $1 - \frac{1}{2N}$, carry genes that are independent in origin from Generation 1, but may have been identical in their origin in Generation 0. The probability the genes are identical by descent from Generation 1 is the inbreeding coefficient of Generation 1 ($F_1 = \frac{1}{2N}$). Note that F_1, F_2, F_t, etc in this section (written in italics) denote inbreeding coefficients, and not generations as on pages 1 and 2 of this section and in earlier sections.

Therefore the probability of identical homozygotes in Generation 2 is:

$$F_2 = \frac{1}{2N} + (1 - \frac{1}{2N})F_1$$

Where F_1 and F_2 are the inbreeding coefficients of Generations 1 and 2. The same arguments apply to future generations, so we can write the recurrence equation:

$$F_t = \frac{1}{2N} + (1 - \frac{1}{2N})F_{t-1}$$

Generation Inbreeding

The inbreeding of any generation is composed of two components: New inbreeding, which arises from self-fertilization and the "old" that was already there.

Note that inbreeding is cumulative and the absence of inbreeding in Generation t does not change the fact that a population had old inbreeding from prior generations.

Through a series of algebraic steps, we can write the inbreeding coefficient as a function of the number of generations removed from the base populations:

$$\mathrm{F}_t = 1 - (1 - \Delta F)^t$$

where,

$$\Delta \mathrm{F} = \frac{1}{2N}$$

Genotype Frequencies

The genotype frequencies in a population can then be expressed as:

Table 1 Genotype frequencies for a locus with two alleles. Data from Falconer ad Mackay, 1996.

	Original Frequencies	Change due to inbreeding	Origin	
			Independent	Identical
AA	p_0^2	$+p_0 q_0 F$	$= p_0^2(1-F)$	$+p_0 F$
Aa	$2p_0 q_0$	$-2p_0 q_0 F$	$= 2p_0 q_0(1-F)$	
aa	q_0^2	$+p_0 q_0 F$	$= q_0^2(1-F)$	$+q_0 F$
			Allozygous genes	**Autozygous genes**

We can examine inbreeding or the probability of identity of alleles by descent, by looking at the genotype frequencies in Table 1 under two extremes of the *F* statistic:

- if ***F* = 0** (random mating; no inbreeding)
  - the equation for genotypic frequencies reduces to the familiar equation for genotypes in Hardy-Weinberg proportions: $p^2 + 2pq + q^2$
- if ***F* = 1** (alleles identical by descent)
  - the genotype equation reduces to a ratio of homozygotes to heterozygotes as *p:0:q*

Inbreeding Coefficient

Thus inbreeding leads to homozygosity (all or nearly all loci homozygous), and almost a complete absence of heterozygosity (all or nearly all loci heterozygosity). As noted in Table 1, with inbreeding there is a deficit of heterozygotes equal to 2pq*F* and an excess of each homozygous class equal to half the deficiency of the heterozygotes.

To illustrate one type of mating that promotes inbreeding, consider the case of a population that reproduces by self-fertilization so that *F* = 1.

Another way to express the inbreeding coefficient, *F*, is to compare the frequency of heterozygotes in the population to the frequency expected under random mating:

$$F = \frac{H_0 - H}{H_0}$$

where,

- **H** = frequency of heterozygotes in the population
- $\mathbf{H_0}$ = expected frequency under HWE, meaning 2pq

Therefore the inbreeding coefficient is the proportional reduction in heterozygosity relative to a random mating population with the same allele frequencies.

Inbreeding via Self-Pollination

In a population that reproduces by self-fertilization the inbreeding coefficient, $F = 1$. Let's assume the population begins with genotypic frequencies in Hardy-Weinberg proportions ($p^2 + 2pq + q^2$). With selfing, each homozygote produces only progeny of the same genotype:

AA × AA ⇒ all AA

aa × aa ⇒ all aa

However only half of the progeny of a heterozygote will be like the parent

Aa × Aa ⇒ 1/4 AA, 1/2 Aa, 1/4 aa

Self-pollination therefore reduces the proportion of heterozygotes in the population by half with each generation until all genotypes in the population are homozygous (Table 2).

Table 2 Increased over generations in frequency of homozygous in a self-pollinated population starting with p = q = 0.05>>>frequency of homozygotes.

	Genotype Frequencies		
Generation	**AA**	**Aa**	**aa**
1	1/4	1/2	1/4
2	1/4 + 1/8 = 3/8	1/4	1/4 + 1/8 = 3/8
3	3/8 + 1/16 = 7/16	1/8	3/8 + 1/16 = 7/16
4	7/16 + 1/32 = 15/32	1/16	7/16 + 1/32 = 15/32
n	$\frac{1-(1/2)^n}{2}$	$(1/2)^n$	$\frac{1-(1/2)^n}{2}$
	1/2	0	1/2

Proportion of Homozygotes

With multiple segregating loci, the proportion of homozygotes in various selfing generations can be estimated using the following formula: $[(2^m - 1)/2^m]^n$: where m is the number of selfing generations(m = 1 for F_2; m = 2 for F_3 and so on) and n is the number of segregating loci. For example, an F_1 plant with four independent segregating loci will result in the following frequency of homozygous plants in F_2, $[(2^1 - 1)/2^1]^4 = 1/16 = 6.25\%$. The expected proportion of completely homozygous plants in F_2 and later generations of selfing for different numbers of segregating loci are indicated in Table 3. The proportion of homozygotes decreases sharply with increasing heterozygosity (more segregating loci) in the F_1.

Table 3 Frequency of completely homozygous individuals in various selfing generations in relation to the number of segregating loci in F_1.

Number of segregating loci in the F_1	Frequency of completely homozygous individuals (%)			
	F_2	F_4	F_6	F_8
1	50.00	87.5	96.87	99.22
2	25.00	76.56	93.85	98.44
3	12.50	66.99	90.91	97.67
4	6.25	58.62	88.07	86.91
5	3.13	51.29	85.32	96.15
10	0.01	26.31	72.79	92.45
100	7.89×10^{-31}	6.1×10^{-4}	4.18	45.64
1000	9.33×10^{-302}	1.02×10^{-58}	1.63×10^{-58}	0.04

Consequences of Inbreeding

Increasing Inbreeding

Through increasing homozygosity, inbreeding brings together identical alleles at a locus. Homozygosity permits the expression of recessive alleles that may have been previously masked by heterozygosity in the parent generation. If recessive alleles are less favorable than dominant ones, the overall fitness of the individual decreases. Inbreeding is often detrimental because it increases the appearance of lethal and deleterious recessive traits. The term

inbreeding depression describes the decrease in fitness or performance that often accompanies inbreeding or random genetic drift. Recall that fitness is the relative ability of an individual to survive and reproduce to contribute its genes to the next generation. Inbreeding depression is further described in the next section.

Plant stature, vigor, yield and other traits decline with increasing inbreeding, although significant differences exist among species for the amount of inbreeding depression expressed—ranging from minimal among self-pollinated crops such as oat and wheat to severe in cross-pollinated polyploid species such as alfalfa, whereby homozygous genotypes do not survive (Fig. 4). Table 4 on the next screen summarizes some general differences between outbreeders and inbreeders.

Fig. 4 The effect of inbreeding on plant stature, vigor, yield and other traits.

Outbreeders vs. Inbreeders

Table 4 General features of reproduction and population genetics pertaining to outbreeders vs. inbreeders. Data from Simmonds and Smartt, 1999.

Outbreeder	Inbreeder
Has crossing mechanism, approaches random mating	Closed flowering, approaches regular selfing
Individuals heterozygous at many loci	Individuals approach homozygosity
Variability distributed over the population	Variability mostly between component lines
Carries deleterious recessives	Deleterious recessives tend to be eliminated
Intolerant of inbreeding	Tolerant of inbreeding
Much heterozygote advantage (epistasis, overdominance)	Less heterozygote advantage

Inbreeding Depression

Inbreeding Depression and Homozygosity

Inbreeding results in increased homozygosity. In cross-pollinated species, increased homozygosity results in inbreeding depression or reduced performance. Symptoms of inbreeding depression may include:

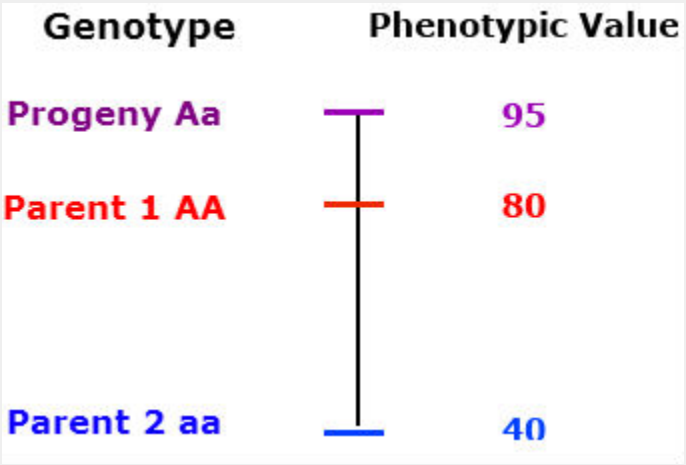
- Reduced plant vigor
- Smaller plant size
- Decline in fertility
- Suppressed seed production
- Decreased pollen production
- Inferior seed quality
- Greater susceptibility to insect or pathogen damage or
- Poorer standability.

Why? With increased homozygosity, expression of deleterious recessive alleles may be revealed by fixation that had been masked by more favorable dominant alleles, or it can be caused by overdominance in which homozygotes are less fit, resulting in poorer performance.

The severity of inbreeding depression varies with the species and genotype. If the inbreeding depression is too severe, it may be difficult or impractical to maintain or propagate inbred lines by seed. In such cases, some heterogeneity must be maintained or other propagation means used.

Overdominance Hypothesis

Over-dominance: The phenotype of the heterozygous progeny is greater than either parent.



Mating Systems

The frequency of homozygotes in a population can be increased using any of several mating systems:

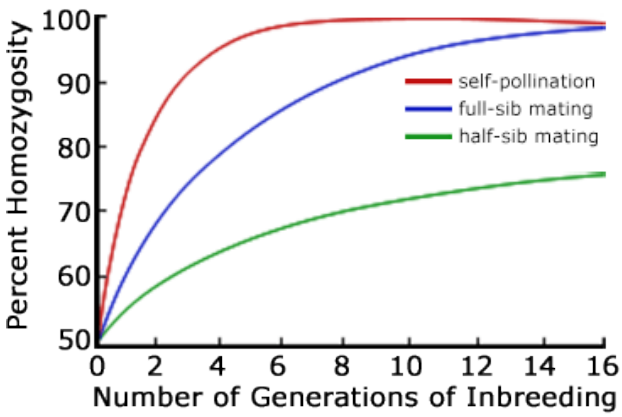


Fig. 5 Comparison of homozygosity reached by self-pollination, full-sib mating, and half-sib mating.

- **Self-pollination**—Repeated self-pollination is routinely used to develop pure lines of self-pollinating crops. It can also be applied to cross-pollinating species to obtain inbred lines. Like pure-lines in self-pollinated species, inbred lines are homozygous at nearly all loci.
- **Half-sib mating**—Plants having **one parent in common** are mated. The pollen source is random from the breeding population, but the female plants are identifiable.
- **Full-sib mating**—Plants having **both parents in common** are crossed.

- **Backcrossing**—A method in which a hybrid is mated to one of its parents, the recurrent parent, resulting in a backcross 1 (BC_1) population. The resulting backcross offspring are repeatedly crossed to the recurrent parent.

The inbreeding coefficient increases more rapidly with matings among more closely related individuals. The first three mating systems listed above are compared in Fig. 5. In each case the initial value of the inbreeding coefficient is assumed to be $F_0 = 0$. Backcrossing—like self-fertilization—can be one of the most extreme forms of inbreeding.

Degree of Relatedness of Individuals

The coefficient of relatedness (r_{IJ}) is a measure of the degree of relatedness between individuals (Table 5). It is the proportion of genes shared between two individuals *I* and *J* due to common descent (e.g., identical by descent). It ranges from -1.0 (no genes in common, at least among the genetic markers assessed) to +1.0 (vegetative clones). For example, the value of r_{IJ} for a parent and each of their progeny is 0.5 for a sexually reproducing species because half of each offspring's genes come from each parent. In the case of full-siblings the value of r_{IJ} is also 0.5, but on average because siblings receive half of their genes from each of the same pair of parents, but the haploid set of genes in each parental gamete is a random sample of half of the parental genome due to recombination (Conner and Hartl 2004).

Table 5 Values of the coefficient of relatedness (r_{IJ}) among relatives. Data from Conner and Hartl, 2004.

Mating Systems	**r_{IJ}**
Self-pollination*, vegetative clones, doubled haploids	1.0
Parents and offspring	0.5
Full siblings	0.5
Half siblings	0.25

*After selfing for many generations

Heterosis

Heterosis or Hybrid Vigor

When inbred lines of cross-pollinating species are mated, their progeny are often superior to either or both of the parents for one or more characters, a phenomenon referred to as **hybrid vigor** or **heterosis**. Recall that genetic drift is random, so that different inbred lines or subpopulations tend to be fixed for different alleles at various loci; when they are intermated, the F_1 population will be highly heterozygous. Heterosis is the opposite of inbreeding depression—heterosis is commonly expressed as

- Improved plant vigor,
- Greater plant size, and/or
- Increased productivity.

The performance of a hybrid relative to its parents can be described in two ways: mid-parent heterosis is the performance of a hybrid compared with the average of the performance of its parents; high-parent heterosis is a comparison of the performance of the hybrid with that of the best parent:

$$\text{mid-point heterosis } (\%) = \frac{F_1 - MP}{MP} \times 100$$

Where:

- F_1 = performance of the hybrid
- MP = average performance of the parents (Parent 1 + Parent 2)/2
- HP = performance of the best parent

Dominance vs. Overdominance Hypothesis

Heterosis tends to be greatest in the progeny of diverse genotypes. There are two common explanations for this phenomenon:

1. Dominance hypothesis—Heterosis results from the complete or partial dominance of favorable alleles at various loci.

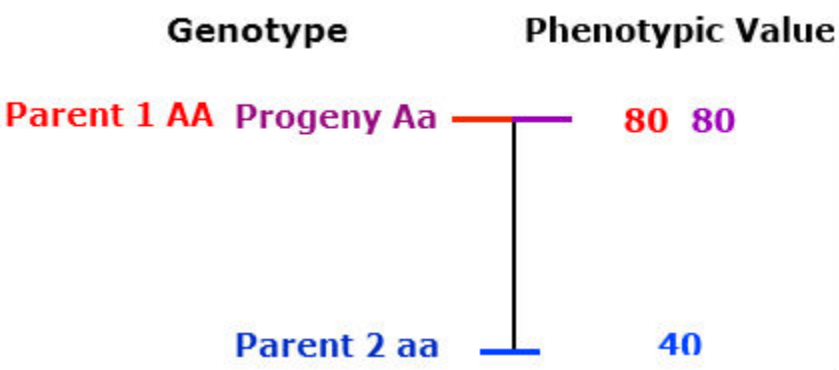
2. Overdominance hypothesis—Heterosis reflects the superior performance of the heterozygote over either homozygote.

The next few pages provide a summary of the dominance and overdominance hypotheses from Charlesworth and Willis (2009). Note the differences between models for a single locus and those describing multiple loci. They include a third model that involves deleterious alleles at closely linked loci and focuses on so-called "pseudo-overdominance", which will not be further discussed here.

Dominance Hypothesis

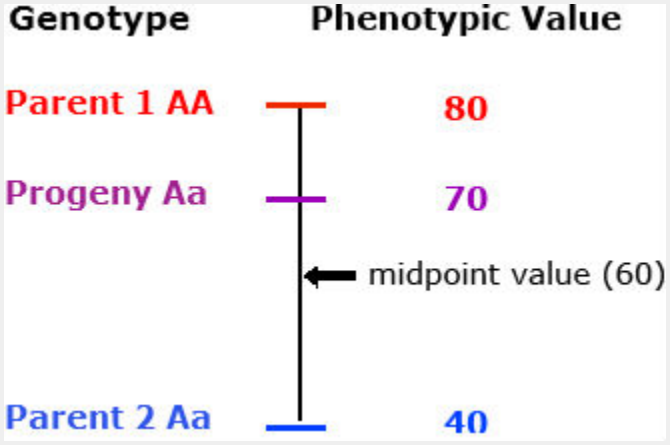
Complete dominance

The phenotype of the heterozygous progeny equals the phenotype of the homozygous dominant parent.



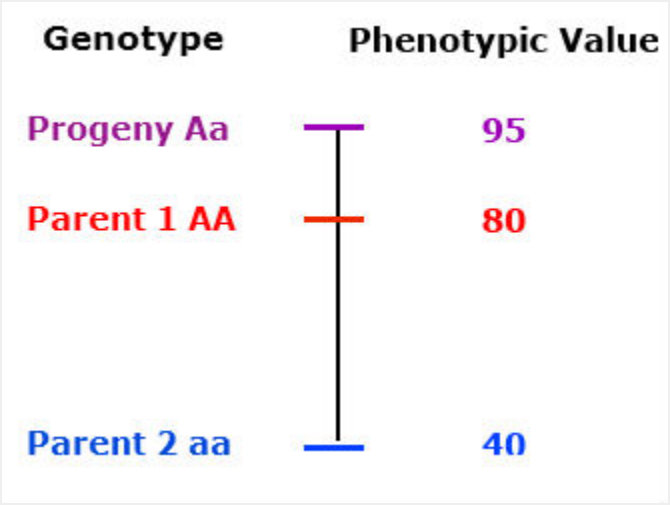
Partial (incomplete) dominance

The heterozygous progeny has a phenotypic value greater than that of the midparent value, but less than that of the homozygous dominant parent.



Overdominance

The phenotype of the heterozygous progeny is greater than either parent.



Main Genetic Hypothesis

Recessive deleterious mutations

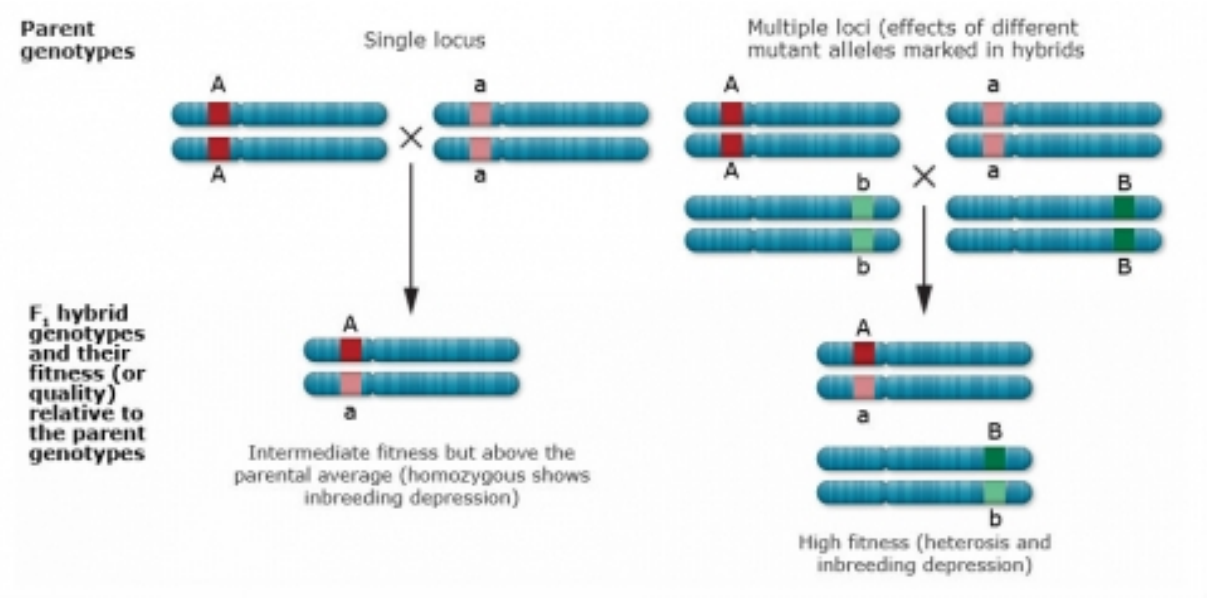


Fig. 6 Dominance hypothesis. Summary of the main genetic hypotheses for inbreeding depression and heterosis. Adapted from Charlesworth and Willis, 2009.

Recessive deleterious mutations at closely linked loci

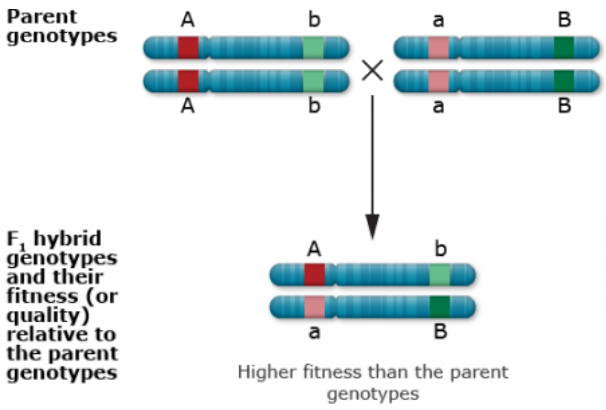


Fig. 7 Pseudo-overdominance. Summary of the main genetic hypotheses for inbreeding depression and heterosis. Adapted from Charlesworth and Willis, 2009

Single locus with heterozygous advantage

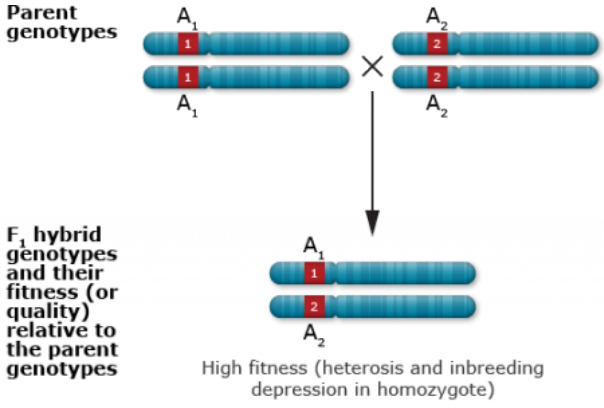


Fig. 8 True overdominance. Summary of the main genetic hypotheses for inbreeding depression and heterosis. Adapted from Charlesworth and Willis, 2009.

Heterozygote Advantage

Heterozygote advantage is a term that is frequently confused with heterosis. However, heterozygote advantage is a synonym of overdominance, and refers to a condition in which the heterozygous genotype has a higher phenotypic value (especially for fitness) than for either homozygous genotype. Heterosis is usually due to a number of loci that control quantitatively inherited traits, although depending on the situation it could be due to phenotypes controlled by either single genes or polygenic traits. In contrast, heterozygote advantage describes effects at a single locus (Conner and Hartl 2004). Compare the circumstances for these two concepts in Table 6.

Table 6 Fitness effects of high and low frequency of heterozygotes. Data from Donner and Hartl, 2004.

	Subpopulations	
	Small and isolated	**Crossed with each other**
Genotypic frequencies	Highly homozygous	Highly heterozygous
Fitness	Low	High
Fitness effects	Inbreeding depression	Heterosis or hybrid vigor
Cause for fitness effects	Deleterious recessives expressed or loss of heterozygote advantage	Deleterious recessives masked or occurrence of heterozygote advantage

Genetics of Heterosis and Inbreeding Depression

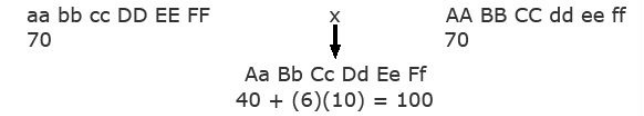
Whatever genetic mechanisms can explain heterosis must also be able to explain inbreeding depression. Let's take a look at hypothetical numeric examples of these two hypotheses to show how each can explain both.

Heterosis and Inbreeding Depression

In the following examples we will use a scenario where six genes control a certain quantitative characteristic.

Favorable Dominant Alleles Theory (with complete dominance at each locus)

1. **Heterosis** The homozygous recessive genotype, **aa bb cc dd ee ff** has a value of 40. One dominant allele at any of the six genes increases the value by 10, that is, *N_*=+10.The following two homozygous parents are crossed. Each has a value of 70 because of having three loci with dominant alleles [40+(3)(10)=70].



Note that the F_1 has a value of 100 because of having six loci with dominant alleles (40 + (6)(10) = 100).

2. **Inbreeding depression**Inbreeding depression can be explained with the accumulation of favorable dominant alleles mechanism. If we self the **Aa Bb Cc Dd Ee Ff** individuals that have a value of 100, we will get a population of individuals as follows:

Table 7

Percentage of individuals	Number of loci with at least one dominant allele
17.80%	6
35.60%	5
29.66%	4
13.18%	3
3.30%	2
0.44%	1
0.02%	0

This gives an overall population mean value of 85. This is lower than the value of the **Aa Bb Cc Dd Ee Ff** parental individuals, thus showing depression upon inbreeding.

Although conclusive evidence to support either of these hypotheses remains elusive, the dominance hypothesis has been more widely accepted. However, it is recognized that a gene's effect is determined by its effect in combination with:

- Other alleles at the same locus,
- Itself (when homozygous),
- Genes at other loci (epistasis), and
- Closely linked genes.

Thus, dominance, overdominance, and epistasis, most likely all contribute to heterosis.

Overdominance Theory

1. **Heterosis:** The genotype **aa bb cc dd ee ff** has a value of 40.The homozygous dominant allele at any locus increases the value by 7 over the homozygous recessive allele, i.e., **NN** = +7. Heterozygous at any locus increases the value by 10 over the homozygous recessive, i.e. **Nn** = +10. The following homozygous parents are crossed; each has a value of 61 because of having three loci with homozygous alleles [40 + (3)(7) = 61]. The F_1 has a value of 100 because of having six heterozygous loci.
2. **Inbreeding depression:** Inbreeding depression can be explained with overdominance. If we self the **Aa Bb Cc Dd Ee Ff** individuals that have a value of 100, we will get a population of individuals with a mean value of 85 (see table above). As was the case for the "favorable dominant alleles theory", this value is lower than the value of the **Aa Bb Cc Dd Ee Ff** parental individuals, thus showing depression upon inbreeding.

Genetic Analysis

Molecular techniques are now being used to try to shed more light on the mechanisms involved in heterosis. In a study published in 2006, Swanson-Wagner et al. assayed the gene expression of 13,999 maize genes from seedling plants of two inbred parents and their F_1 hybrid. When

comparing the three genotypes they found 1,367 of the genes produced significantly different amounts of messenger ribonucleic acid (mRNA), which performs an essential role in protein synthesis.

Analysis of these genes showed that 78 percent of them had additive gene action, 15 percent showed either high-parent or low-parent dominance, and 3 percent showed either overdominance or underdominance. (The other 45 genes showed non-additive gene action but with the statistical analysis used these genes could not be classified as dominant, overdominant, or underdominant.) The experiment analyzed individual genes and was not designed to evaluate interactions among the genes, so epistasis could not be measured.

This study involved only two maize inbred lines and their hybrid offspring and the evaluation was based on tissue from seedling plants. Different genes will be active during different parts of the life cycle of the maize plant and further studies will most likely show differential gene action among different genes. Likewise, different parents and hybrids and different species may well show very different gene action patterns among the same genes that were tested in this study. With just this one experiment, where nearly 14,000 genes were evaluated, we can see that heterosis is complex and that multiple types of gene action are involved. Molecular methods in the context of hybrid breeding will be discussed in detail in the molecular plant breeding course.

When two genetically different parents are mated, heterosis is observed in their F_1 progeny.

- All seed and forage species express some level of heterosis.
- Most cross-pollinated species express greater heterosis than do self-pollinated species.

Because of the expense in producing hybrid cultivars, hybrid cultivars are generally developed only in those crops in which the level of heterosis results in significantly better performance.

References

Charlesworth, D. and J.H. Willis. 2009. The genetics of inbreeding depression. Nature Reviews Genetics 10: 783-796.

Conner, J. K., and D.L. Hartl. 2004. A Primer of Ecological Genetics. Sinauer Associates, Sunderland, MA.

Falconer, D.S. and T.F.C. Mackay. 1996. Introduction to Quantitative Genetics. 4th edition. Longman Pub. Group, Essex, England.

Pierce, B.A. 2010. Genetics: A Conceptual Approach. 3rd edition. W.H. Freeman, New York, NY.

Simmonds, N.W. and J. Smartt. 1999. Evolution of Crop Plants. 2nd edition. Longman, Essex, UK.

Swanson-Wager, R.A., Y. Jia, R. DeCook, L.A. Borsuk, D. Nettleton, and P.S. Schnable. 2006. All possible modes of gene action are observed in a global comparison of gene expression in a maize F_1 hybrid and its inbred parents. PNAS 103:6805-6810.

How to cite this chapter: Merrick, L., W. Beavis, J. Edwards, T. Lübberstedt, A. Campbell, D. Muenchrath, and S. Fei. (2023). Inbreeding and heterosis. In W. P. Suza, & K. R. Lamkey (Eds.), *Crop Genetics*. Iowa State University Digital Press. DOI: 10.31274/isudp.2023.130

Inheritance of Quantitative Traits

Laura Merrick; Kendra Meade; Arden Campbell; Deborah Muenchrath; Shui-Zhang Fei; and William Beavis

Introduction

Many of the traits that plant breeders strive to improve are quantitatively inherited. For example, breeding efforts targeting quantitative traits have allowed major increases in crop yield during the past 80 years or so. A quantitatively inherited trait is controlled by many genes at different loci, with each gene — known as a **polygene** — contributing a small effect to the expression of the character. Polygenes are also known **quantitative trait loci** (**QTL**). QTLs involved in expression of a quantitative character act cumulatively to determine the phenotype of the trait. Their mode of inheritance is called quantitative genetics. Quantitative genetics describes the connection between phenotype and genotype and provides tools to show how phenotypic selection of complex characters changes allele frequencies.

Quantitative genetics focuses on the nature of genetic differences, seeks to determine the relative importance of genetic vs. environmental factors, and examines how phenotypic variation relates to evolutionary change. Typically, quantitative genetic analysis is executed on traits showing a continuous range of values. Analysis of quantitative traits is based on statistical predictions of population response. Examples of quantitatively inherited traits include yield, vigor, rate of photosynthesis, protein content, and drought tolerance.

Key Concepts

Central to the mathematical modeling of quantitative genetics is the concept of recognition of family resemblance. If genes influence variation in a trait (and sources of environmental variation are minimized or controlled), related individuals would be expected to resemble one another **more than** unrelated ones. Siblings should resemble each other **more than** distantly related relatives. A comparison of plant individuals with different degrees of relatedness provides information about how much genes influence the character.

Several factors influence the likelihood of progress when breeding for quantitatively inherited traits, including:

- Interaction of the multiple genes contributing to the phenotype;
- Gene actions of the respective genes involved; and
- Frequencies of those genes.

Various models are used to distinguish genetic, environmental, and genetic x environmental interaction effects on phenotype and to improve breeding efficiency of quantitative traits.

Learning Objectives

- Distinguish environmental and hereditary variation and be aware of why detecting the interaction of genetic and environmental factors is important.
- Understand the attributes of quantitative inheritance.
- Be familiar with statistical methods applied to quantitative inheritance.
- Study the types of gene actions and interactions affecting quantitative traits.
- Examine the genetic advance from selection formula and be able to explain how each of its components influences the improvement of the selected characteristic.
- Learn the types of heritability estimates and their importance in plant breeding.

Heritable vs. Environmental Variation

The **phenotype** of a plant or group of plants is modeled as a function of its **genotype** as modified by the environment.

Phenotype = Genotype + Environment + (Genotype x Environment)

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

Some characters are **more responsive or sensitive to growing conditions than others**. Qualitative traits such as flower color are not strongly influenced by the environment. On the other hand, quantitative traits such as grain yield or abiotic stress tolerance are influenced markedly by the environment. The degree of sensitivity or the range of potential responses

to the environment is determined by the genetic composition of the individual plant or population of plants.

Genetic variation is essential in order to make progress in cultivar improvement. However, sources of variation include:

- Environmental variation
- Genetic variation and
- Interaction of genetic and environmental variation

Plant breeders must distinguish among these sources of variation for the character of interest in order to effectively select and transmit the desired character or assemblage of characters to subsequent generations.

GxE Interaction Example

A classic example of genotype × environment interaction (GxE) involves studies conducted in the 1930s and 1940s by the ecologists Clausen, Keck, and Hiesey (1940, 1948). They collected plants from natural wild populations — principally yarrow (Achillea millefolia) and sticky cinquefoil (Potentilla glandulosa) — that grew along an east-west elevational transect in California, running from near sea level at the Pacific Ocean to more than 3000 m elevation in the Sierra Nevada mountains.

Both species exhibited a vast amount of variation in native populations with respect to growth form and other traits, including such attributes as plant height, winter survival, and number of stems produced. Through a series of reciprocal transplant experiments using cuttings or clonal material from wild populations, they tested the contribution of genetic and environmental variation to observed phenotypic variation among plants established in three main “common gardens” or transplant plots: Stanford, Mather, and Timberline (Fig. 1).

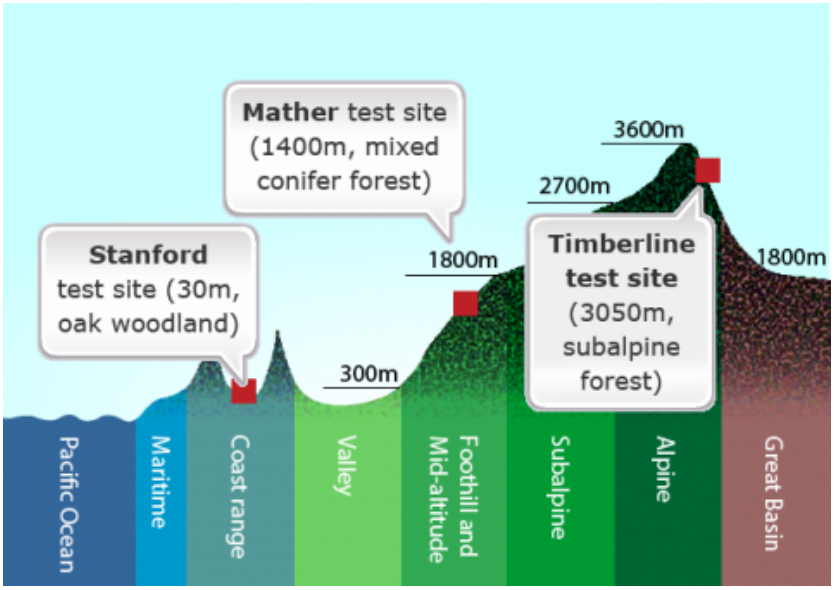


Fig. 1 Locations for experimental studies examining contributions of G, E, and GxE. Adapted from Barbour at al., 1999.

Conclusions

They concluded each species had differentiated into genetically distinct subspecies — which they called ecotypes — that are best suited to their specific environments. In the transplant gardens, no single ecotype performed best at all altitudes. For example, genotypes that produced the tallest plants at the mid-altitude garden site grew poorly at the low and high sites. Conversely, genotypes that grew the best at the low or high sites sometimes performed poorly at the mid-altitude site. Although within a species, all populations were found to be completely interfertile, ecotypes adapted to low or mid-altitude died when transplanted to the high altitude garden, while ecotypes from high elevations along the transect survived through the winter when locally grown in the test plots. GxE interaction was observed for height, among other characters.

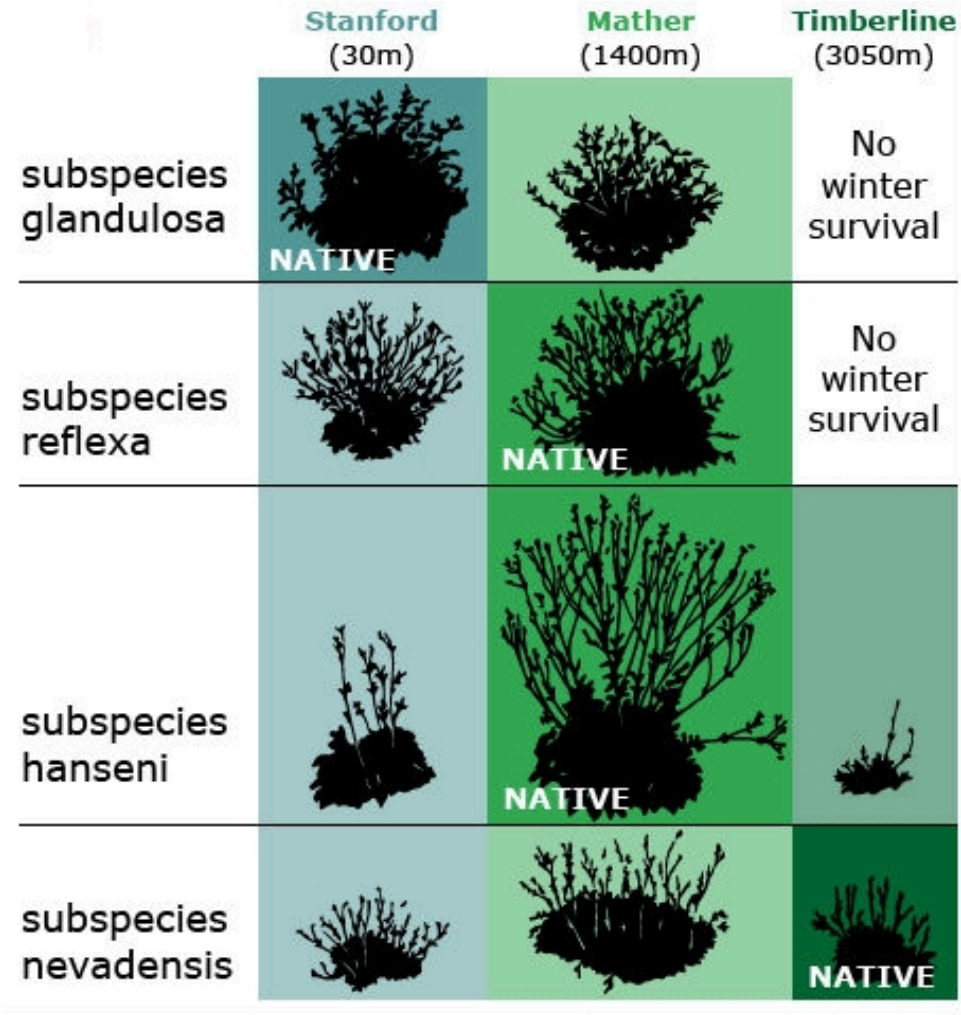


Fig. 2 Response of plant ecotypes (Potentilla subspecies) to low, mid-, and high elevation transplant sites. Data from Clausen, Keck, and Hiesey, 1940, in Rauscher, 2005.

Significance Illustration

Figure 3 shows variation in phenotype between two cultivars of watermelon with regard to a quantitatively inherited trait (yield in this case) in response to variation in an environmental factor (soil salinity in this case) and illustrates the significance of genotype x environmental interaction.

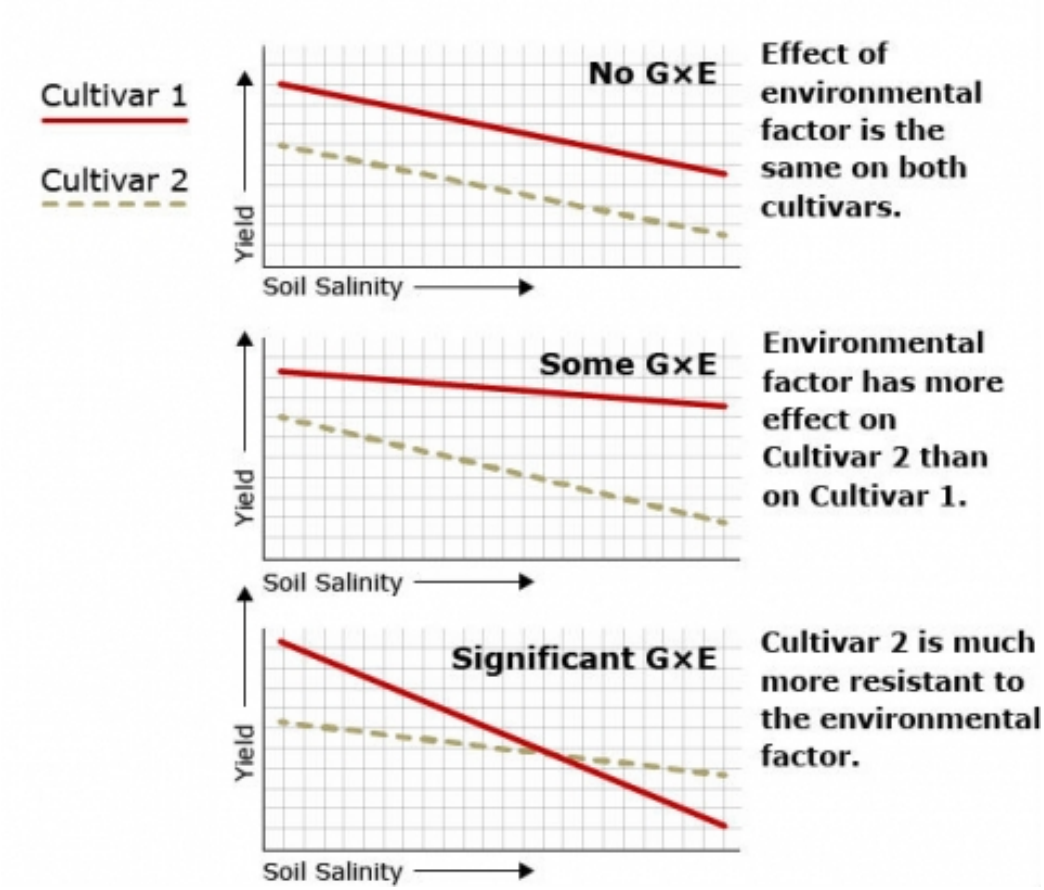


Fig. 3 Hypothetical comparisons of genetic x environment interactions (GxE) in yield of two watermelon cultivars in response to increasing levels of soil salinity.

Characteristics of Quantitative Traits

Inheritance of quantitative traits involves two or more nonallelic genes (multiple genes or polygenes); the combined action of these genes, as influenced by the environment, produces the phenotype. The effect of individual genes on the trait is not apparent. However, early in the 1900s it was discovered that the inheritance of the individual genes contributing to the phenotype of quantitative traits do indeed follow the same Mendelian inheritance principles as simply-inherited genes.

Fig. 4 The diameter of Gaillardia pulchella flowers is a quantitative trait. Photo by DanielCD; CC-SA 3.0 via Wikimedia Commons.

Inheritance of Quantitative Traits

Fig. 5 Herman Nilsson-Ehle. Photo licensed under Public Domain via Wikimedia Commons.

In 1909, Herman Nilsson-Ehle, a Swedish geneticist and wheat breeder, conducted some of the classic studies on quantitatively inherited traits in wheat. He developed what is known as the "Multiple Factor" or multi-factorial theory of genetic transmission. A key observation made by Nilsson-Ehle was that although a spectrum of continuous variation in kernel color (a quantitatively inherited trait influenced by environmental factors) could be observed in segregating generations, he was able to determine that segregation for these genes fit a model that each separate contributing gene followed a pattern of **Mendelian inheritance**.

Fig. 6 Wheat kernels. Photo by zandland; licensed under CC-SA 3.0 via Wikimedia Commons.

Bread wheat is a hexaploid — allopolyploid that contains three slightly different, but similar ancestral genomes (referred to as A, B, and D) in its genome (AABBDD). Depending on the cultivars that Nilsson-Ehle studied, each genome had a single gene that affected kernel color, and each of these loci has a red allele (**R**) and a white allele (**r**). Alleles at each locus varied slightly in their effect on kernel color, and will be designated in this example by different superscripts, e.g., $\mathbf{R^1}$ or $\mathbf{r^3}$.

He crossed two cultivars of wheat that varied in kernel color, one with dark red seeds (homozygous dominant genotype $\mathbf{R^1R^1R^2R^2R^3R^3}$, based on the symbols designating the ancestral genomes) and another with white kernels (homozygous recessive $\mathbf{r^1r^1r^2r^2r^3r^3}$). He noted that the F_1 of a cross between these parents (heterozygote $\mathbf{R^1r^1R^2r^2R^3r^3}$) was intermediate in color (light red), but the F_2 generation could be grouped into seven classes, ranging in color from dark red to white. He explained the distribution on the basis of three pairs of genes segregating independently, with each dominant allele contributing to the intensity of the red color.

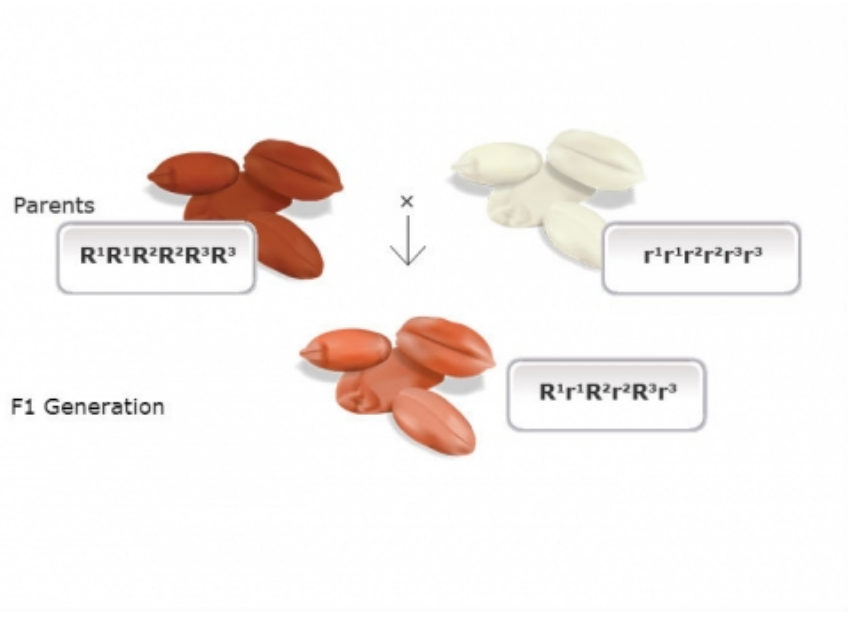


Fig. 7

Table 1 Kernel color in F_2 progenies from a wheat cross.

F_2 genotypes			Color	Number of dominant alleles	Number of plants out of 64
	$R^1R^1R^2R^2R^3R^3$		dark red	6	1
$R^1R^1R^2R^2R^3r^3$	$R^1r^1R^2R^2R^3R^3$	$R^1R^1R^2r^2R^3R^3$	moderately dark red	5	6
$R^1R^1R^2R^2r^3r^3$ $r^1r^1R^2R^2R^3R^3$	$R^1R^1r^2r^2R^3R^3$ $R^1r^1R^2r^2R^3R^3$	$R^1r^1R^2R^2R^3r^3$ $R^1R^1R^2r^2R^3r^3$	red	4	15
$R^1R^1R^2r^2r^3r^3$ $R^1R^1r^2r^2R^3r^3$	$R^1r^1R^2r^2R^3r^3$ $R^1r^1R^2R^2r^3r^3$ $R^1r^1r^2r^2R^3R^3$	$r^1r^1R^2R^2R^3r^3$ $r^1r^1R^2r^2R^3R^3$	light red	3	20
$R^1R^1r^2r^2r^3r^3$ $r^1r^1R^2R^2r^3r^3$	$r^1r^1r^2r^2R^3R^3$ $R^1r^1R^2r^2r^3r^3$	$R^1r^1r^2r^2R^3r^3$ $r^1r^1R^2r^2R^3r^3$	pink	2	15
$R^1r^1r^2r^2r^3r^3$	$r^1r^1R^2r^2r^3r^3$	$r^1r^1r^2r^2R^3r^3$	light pink	1	6
	$r^1r^1r^2r^2r^3r^3$		white	0	1

With three independent pairs of genes segregating, each with two alleles, as well as environmental effects acting on kernel color, the F_2 progeny would contain 63 plants with varying shades of red kernels and one with white kernels. Linkage among the genes restricts independent assortment, so that the required size of the F_2 population becomes larger.

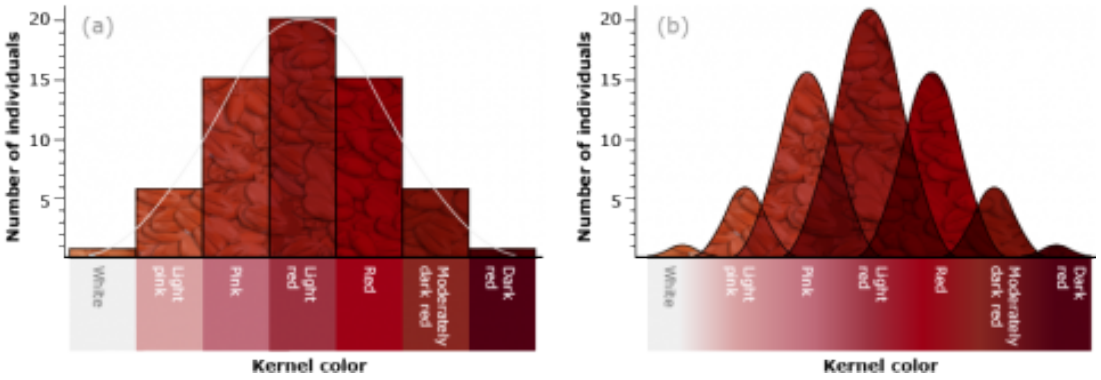


Fig. 8 Range of wheat kernel color in an F_2 generation. (a) Kernel color depicted by seven discrete classes modeled on segregation of three contributing genes, each exhibiting partial dominance. (b) Kernel color depicted by continuous variation in all seven color classes.

Characteristics Indicative of Quantitative Inheritance

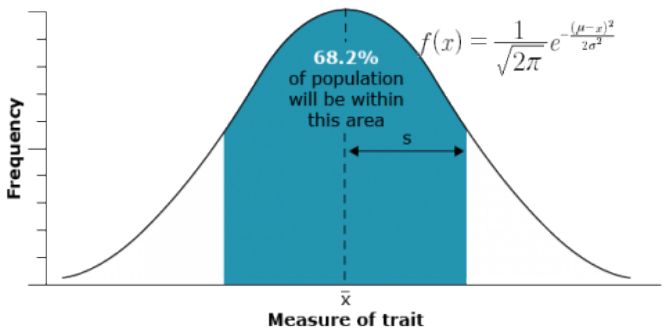


Fig. 9 Based on a random sample from a genetically mixed population, the distribution of a quantitative trait's expression approximates a normal or bell curve.

Phenotypic values for the specific trait, resulting from simultaneous segregation of multiple genes, exhibit continuous variation, rather than distinct classes. In general, the distribution of values of quantitatively inherited traits in a population follows a **normal distribution** (also called a Gaussian distribution or bell curve). These curves are generally characterized by two parameters, the **mean** and the **variance** or **standard deviation**. In the figure below, the mean of the random sample is depicted by the symbol $\overline{x}$ and the standard deviation by the symbol **s**. If the reference is made to a population instead of a sample from a population, the population mean is usually symbolized by **μ** and the population standard deviation by **σ**.

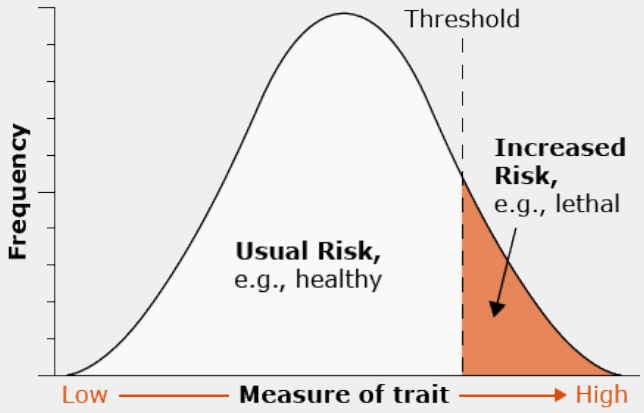
There are three general types of traits that are quantitatively inherited: continuous, meristic, and threshold. An example of the first type, a continuous trait, is fruit width of pineapple. The second type, a meristic character, is a countable trait that can take on integer values only, e.g., number of tillers of maize or branches of a rose bush. The third type of quantitative trait is known as a threshold character or "all-or-none" trait. Such traits are typically ranked simply as presence or absence, e.g., Downy Mildew disease in soybean.

Threshold Traits

Although they have only two phenotypes, threshold traits are considered to be quantitatively inherited because their expression depends on a liability (such as disease susceptibility or tolerance of nicotine levels) that varies continuously. Heritability of these traits is a function of the incidence of the trait in the population, so it is difficult to determine the importance of genetic factors in different environments or in different populations that differ in incidence. Threshold traits are assumed to be represented by an underlying normally distributed "liability

trait" that is the sum of the independent genetic and environmental components of the distribution. A disease would have to be present before you could determine if certain genotypes were susceptible or not. For example, plants might be able to tolerate low to moderate levels of nicotine in their tissues until a threshold was crossed, above which the high level of nicotine present would be lethal.

Threshold characters exhibit only two phenotypes — the trait is either present or absent — but the susceptibility to the trait varies continuously and environmental components of the distribution. A disease would have to be present before you could determine if certain genotypes were susceptible or not. For example, plants might be able to tolerate low to moderate levels of nicotine in their tissues until a threshold was crossed, above which the high level of nicotine present would be lethal.



Threshold characters exhibit only two phenotypes — the trait is either present or absent — but the susceptibility to the trait varies continuously.

Environment has a large influence on the trait's phenotype. That is, for the particular trait, the relative responses of plants change when grown under different environmental conditions.

Distinct segregation ratios of individual nonallelic genes are not observed. Recombination and segregation patterns are based on the **combined** effect of the polygenes on the trait. The more loci controlling the character, the greater the complexity.

Genes may differ in their individual gene action, but their effect on the trait is cumulative. Types of gene action include additive, dominance, overdominance and epistasis. Effects of gene action using the concepts of genotypic and breeding values are discussed in Appendix A.

Genotypic Value

The genotypic value is equal to

$$G = A + D + I$$

- G = genotypic value of all loci considered together
- A = sum of all additive effects (i.e. breeding values) for separate loci
- D = sum of all dominance deviations (i.e., interaction between alleles at a locus or so-called intralocus interactions)
- I = interaction of alleles among loci (also referred to as the deviation or epistatic deviation)

For an individual, the breeding value is calculated by the summation of the average effects of its genes (also referred to as the additive effect of genes). The average effect of an allele is approximately the average deviation of the mean phenotypic value from the population mean if the allele at a particular locus is substituted by another allele (Falconer and Mackay 1996).

Table 2 Summary of interactions among alleles (within or between loci) defining different types of gene action.

	No interaction among alleles	**Interaction among alleles**
Within a locus	Additive	Dominance
Between loci	Additive	Epistasis

Transgressive segregation may occur. These individuals exhibit phenotypes outside the range of those expressed by the parents. Transgressive segregation occurs when progeny contain new combinations of multiple genes with more positive effects or more negative effects for the quantitative trait than found in either parent. One challenge is that strong environmental effects would make it difficult to assess the mean performance in parental plants vs. progeny in order to detect for the presence of any transgressive segregants.

Measurement of Continuous Variation

Analysis of inheritance of qualitative traits is generally concerned with individual matings

and their progeny and is made by counts and ratios. In contrast, analysis of quantitative traits is concerned with populations of organisms that consist of many possible kinds of matings; analysis of such traits is made by use of statistics. Statistical methods provide a tool for describing and evaluating quantitatively inherited characters. Since it is impractical to examine an entire population, plant breeders sample the population(s) of interest. The sample must be representative of the population — the sample must be:

- large enough to include the entire range of variability of the trait that occurs in the population, and
- random to avoid introducing any bias.

Thus, the greater the variability within the population, the larger the sample size that is required to accurately describe that population. (Throughout this and subsequent modules, you can generally assume that we're referring to a representative sample, rather than a population.)

Statistics may be descriptive or analytical. [See Appendix B to briefly review some statistical terms and concepts].

Statistical Terms

The study of quantitative traits is sometimes referred to as "statistical genetics" because of its reliance on statistical methods. In order to understand the inheritance of quantitative characters and the methods applied to these characters, it is essential that you become familiar with fundamental statistics. A basic review is provided here.

When using symbols to represent these population parameters, it is important to distinguish between information about the population and that concerning a sample representing the population.

Similar Statistical Symbols

As a shorthand, statistics commonly use symbols to convey concepts. Often there are several symbols that relate to very similar, but slightly differing concepts. Here's a list of symbols related to means, variance, and standard deviation that you will encounter in this lesson. The

latter two parameters describe the variability or dispersion about the mean of the population or sample derived from a population.

Although the differences between these are important from a statistical perspective, they are commonly used synonymously.

Parameter	For a population	For a fixed or selected sample of a population
Mean	μ or M	$\bar{X}$
Variance	V or δ^2	s^2
Standard deviation	δ	s

Descriptive Statistics

Range — the lowest and highest phenotypic values in the population or sample for the character.

Mean (**μ** or **M** for population; $\bar{X}$ for sample) — describes the average performance of a random sample from a population for a trait. The mean is a measure of central tendency — it does **not** tell anything about the distribution of individual observations. Mean equals the sum of the trait values of each individual divided by the number of samples (n):

$$\bar{x} = \frac{\Sigma x}{n}$$

Variance (**V** or $\boldsymbol{\sigma^2}$ for population; $\mathbf{s^2}$ for sample) — a measure of the scatter or dispersion of phenotypic values. The greater the variability among individuals, the greater the variance. Two populations with the same mean for the same character could differ greatly in their respective variance for that character.

Average the squared deviations from the mean, $(\mathbf{X} - \mathbf{X})^2$:

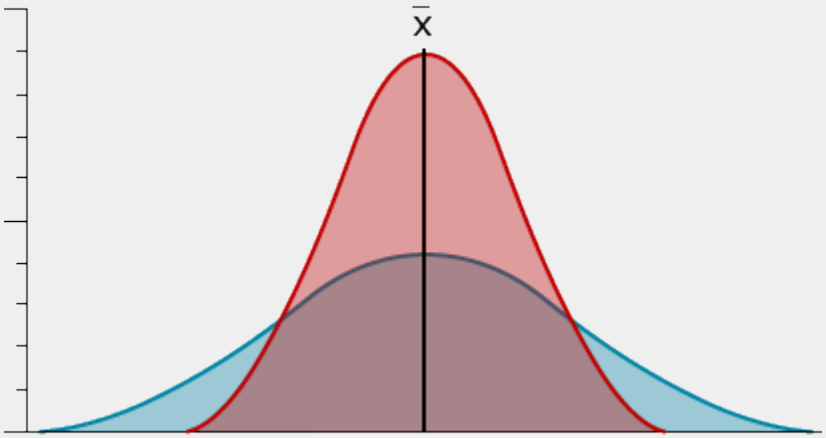
$$V = \sigma^2 = \frac{\Sigma[(x - \bar{x})^2]}{n - 1}$$

Standard deviation (**σ** for population; **s** for sample) — also a measure of dispersion around the mean. Standard deviation expresses the dispersion in the same unit as the mean.

Standard deviation is the square root of the variance.

$$\sigma = \sqrt{V} \text{ or } \sqrt{\sigma^2}$$

A small variance and a small standard deviation tell us that the phenotypic values are near the mean value. In contrast, a large variance and standard deviation indicate that the trait values have a wide range.



Samples in which the observations are clustered closely around the mean (red) have a smaller variance and standard deviation than observations dispersed widely (blue).

Coefficient of variation (CV) — the standard deviation as a percentage of the mean. Because the units cancel out, CV is a unitless measure.

Divide the standard deviation by the mean and multiply by 100:

$$CV = \frac{\sigma}{\bar{X}} \times 100$$

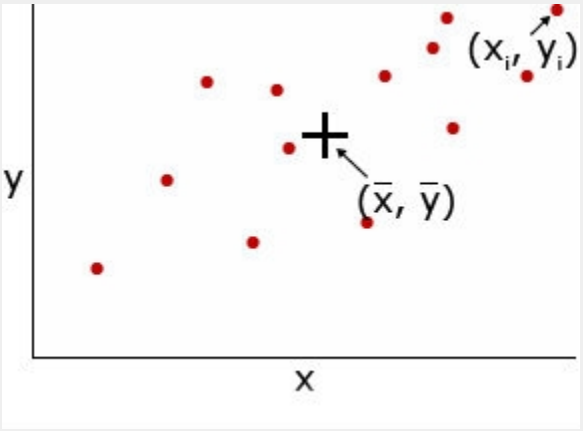
A CV of about 10% or less is desirable in assessing biological systems. When the CV is greater than 10%, the variability in the sample or the population may be too great to sort out the factors contributing to that variability.

The **mean** and the **variance** are used to describe an individual characteristic, but plant breeders are frequently interested in more than one trait simultaneously. Two or more characteristics can vary together, and are thus not independent of one another.

Covariance provides measure of the strength of the **correlation** or dependent relationship between two or more sets of variables. When two traits are correlated, a change in one trait is likely to be associated with a change in the other trait. **Variance** is a special case of covariance that is the covariance of a variable with itself. Correlations between characteristics are measured by a **correlation coefficient** (**r**).

Covariance — the mean value of the product of the deviations of two random variables from their respective means. For example, the covariance of two random variables, x and y is expressed as:

$$Cov_{x,y} = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{n-1}$$



Covariance of variables x and y

Correlation coefficient (**r**) — measures the interdependence of two or more variables and is obtained by dividing the covariance of x and y by the product of the standard deviations of x and y. The correlation coefficient can range from -1 to +1.

$$r = \frac{Cov_{x,y}}{S_x S_y}$$

Correlation is a "scaled" version of covariance and the two parameters always have the same sign—positive, negative, or zero (0). When the sign is positive, the variables are positively correlated; when it's negative, they are said to be negatively correlated, and when it's zero, the variable are described as being uncorrelated. But it is important to note that a correlation between variables indicates that they are associated, but it does not imply a cause-and-effect relationship.

QTLs and Mapping

The relative importance of genetic and environmental factors for a given trait can be estimated by the phenotypic resemblance between relatives. Until recently, quantitative genetics focused

on phenotypic information, but increasingly molecular biology tools are being applied in an effort to locate where **quantitative trait loci** (**QTLs**) occur in plant and animal genomes. Various genetic markers have been identified and mapped, allowing **identification of QTLs by linkage analysis**.

A common method for mapping QTLs is to cross two homozygous lines that have different alleles at many loci. The F_1 progeny are then backcrossed and intercrossed to allow genes to recombine through independent assortment and crossing over. Offspring in segregating generations are examined for correlations between inheritance of marker alleles and phenotypes that are quantitatively inherited. QTL mapping will be covered in more detail in later courses on **Molecular Genetics and Biotechnology** and **Molecular Plant Breeding**.

Multiple Genes and Gene Action

The general types of gene action for quantitative characters (additive, full and partial dominance, and over-dominance) do not differ from those for qualitative traits. However, the genes contributing to the phenotype of a quantitative character may or may not differ in their individual gene action, and their relative effects on the trait's expression may differ. Some may have major influence and others may have only minor effects on the phenotype. The genes controlling a quantitative trait may also interact. The table below gives examples of types of gene action at two loci. For quantitative traits this would be expanded to multiple loci.

Table 3

Gene Action or Interaction	Explanation	Example		
Additive Effects	Genes affecting a genetic trait in a manner that each enhances the expression of the trait.	aabb = 0	Aabb = 1	aaBb = 1
		AAbb = 2	AaBb = 2	aaBB = 2
		AABb = 3	AaBB = 3	AABB = 4
Dominance Effects	Deviations from additivity so the heterozygote is more like one parent than the other. With complete dominance, the heterozygote and the homozygote have equal effects	aabb = 0	Aabb = 2	aaBb = 2
		AAbb = 2	AaBb = 4	aaBB = 2
		AABb = 4	AaBB = 4	AABB = 4
Interaction of Epistasis Effects	Two nonallelic genes (e.g., genes at different loci) may have no effect individually, yet have an effect when combined	aabb = 0	Aabb = 0	aaBb = 0
		AAbb = 0	AaBb = 4	aaBB = 0
		AABb = 4	AaBB = 4	AABB = 4
Overdominance Effects	Each allele contributes a separate effect and the combined alleles contribute an effect greater than that of either allele separately	aabb = 0	Aabb = 2	aaBb = 2
		AAbb = 1	AaBb = 4	aaBB = 1
		AABb = 3	AaBB = 3	AABB = 2

In the preceding examples, A and B are assumed to have equal effects. However, this often may not be true because genes at different loci may affect the expression of the trait in different ways. Some QTLs may be genes with major effect, while others may contribute only a minor effect. Penetrance or expressivity (refer to Deviations from Expected Phenotypes in the "Gene Segregation and Genetic Recombination" module) may influence trait expression. Likewise, pleiotropic effects (refer to Gene Interactions in the "Gene Segregation and Genetic Recombination" module) may be present, affecting different traits in different ways.

Fig. 10 Albinism, such as that exhibited by this deer, is one common effect of pleiotropic mutation. Photo by Paulo Brandao; CC-SA 2.0 via Wikimedia Commons.

Heritability

Conceptual Basis for Understanding Heritability

Heritability estimates the relative contribution of genetic factors to the phenotypic variability observed in a population. What causes variance among plants and among lines or varieties? Phenotypic variation observed among plants or varieties is due to differences in

- their genetic makeup,
- environmental influences on each plant or genotype, and
- interaction of the genotype and environment.

The effectiveness with which selection can be expected to take advantage of variability depends on how much of that variability results from genetic differences. Why? Only genetic effects can be transmitted to progeny. Heritability estimates

- the degree of similarity between parent and progeny for a particular trait, and
- the effectiveness with which selection can be expected to take advantage of genetic variability.

Family Resemblance

As mentioned in the introduction of this lesson, central to the understanding of quantitatively inherited traits is the recognition of family resemblance. Two relatives, such as a parent and its offspring, two full or half-siblings, or identical twins, would be expected to be phenotypically more similar to each other than either is to a random individual from a population. Although close relatives may share not only genes (they may also share similar environments for traits that have a large genetic component), resemblance between relatives is expected to increase as closer pairs of relatives are examined because they share more and more genes in common. In this conceptual framework, heritability can be understood as a measure of the extent to which genetic differences in individuals contribute to differences in observed traits.

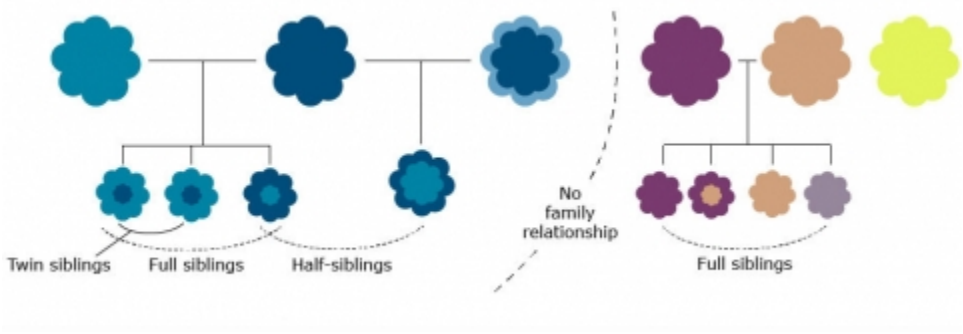


Fig. 11 Individuals that are related genetically would be expected to be more phenotypically similar to each other than to other individuals from a population.

Statistical Basis for Understanding Heritability

For plant breeders, heritability can also be understood in a statistical framework by defining it as the proportion of the phenotypic variance that is explained by genetic variance. Heritability indicates the proportion of the total phenotypic variance attributable to genetic effects, the portion of the variance that is transmittable to offspring. A general formula for calculating heritability is

$$\text{Heritability} = \frac{V_G}{V_P} = \frac{V_G}{V_G + V_E + V_{GE}}$$

$$= \frac{\sigma_g^2}{\sigma_{ph}^2} = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_c^2 + \sigma_{ge}^2}$$

where:

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

or

$$\sigma^2_{ph} = \sigma^2_g + \sigma^2_e + \sigma^2_{ge}$$

and

$V_P = \sigma^2_{ph}$ = phenotypic variance (total variance of the population)

$V_G = \sigma^2_g$ = genotypic variance (variance due to genetic factors)

$V_E = \sigma^2_e$ = environmental variance (variance due to environmental factors)

$V_{GE} = \sigma^2_{ge}$ = genotype x environmental variance (variance due to interactions between genotypes and environmental factors)

Uses of Heritability Estimates

Heritability serves as a guide for making breeding decisions. It is generally used to

- determine the relative importance of genetic effects which could be transferred from parent to offspring
- determine which selection method would be most likely to improve the character
- predict genetic advance from selection

A key point to understand is that heritability is a **population** concept — application of heritability estimates is restricted to the population on which the estimate was based and to the environment in which the population was grown. However, some characters exhibit fairly consistent estimates (either high or low) among populations (within species) and environments. When considering characters that have high heritability, what we expect to observe for each genotype is that its phenotype will be quite predictable over a range of environments (growing conditions). In other words, for characters with high heritability, genotype fairly accurately predicts phenotype. This is not so for characters with low heritability.

Characters

Heritability depends on the range of typical environments experienced by the population under study (if the environment is fairly uniform, then heritability can be high, but if the range of environmental differences is high, then heritability may be low. Even when heritability is high, environmental factors may influence a characteristic. Heritability does not indicate anything about the degree to which genes determine a trait; instead it indicates the degree to which genes determine variation in a trait.

Characters having low heritabilities are usually highly sensitive to the environment, presenting greater breeding challenges — low heritability traits often require larger populations and more test environments than do characters having high heritabilities for selection and improvement.

Table 4 Average heritability estimates (h^2) of maize characters. Average estimates are derived from estimates reported in the literature. The magnitude of these estimates reflects both the complexity of the trait and the number of estimates reported in the literature. Data from Hallauer and Miranda, 1988, p. 118.

Heritability Estimate	Maize Characters
$h^2 < .70$	• Percent Oil • Number of tillers
$.50 < h^2 < .70$	• Plant height • Ear height • Kernel-row number • Dates to flowering • Grain moisture
$.30 < h^2 < .50$	• Number of ears • Ear length • Ear diameter • Kernel weight • Cob diameter
$h^2 < .30$	• Yield • Kernel depth

Broad-Sense Heritability

Types of Heritability

There are two types of heritability: broad-sense and narrow-sense heritability.

Broad-Sense Heritability

Broad sense heritability, H^2, estimates heritability on the basis of all genetic effects.

$$H^2 = \frac{V_G}{V_P} \times 100$$

$$= \frac{\sigma^2 g}{\sigma^2_{ph}} \times 100$$

It expresses total genetic variance as a percentage, and does not separate the components of genetic variance such as additive, dominance, and epistatic effects.

Table 5

Genetic variance	=	Additive variance	+	Dominance variance	+	Epistatic variance
V_G	=	V_A	+	V_D	+	V_I
σ^2_e	=	σ^2_A	+	σ^2_D	+	σ^2_I

Generally, broad-sense heritability is a relatively poor predictor of potential genetic gain or breeding progress. Its usefulness depends on the particular population. Broad-sense heritability is

- more commonly used with asexually propagated crops than with sexually propagated agronomic crops
- applied to early generations of self-pollinated crops

Narrow-Sense Heritability

Narrow-sense heritability, h^2, in contrast, expresses the percentage of genetic variance that is caused by additive gene action, V_A.

$$h^2 = \frac{V_A}{V_P} \times 100$$

$$= \frac{\sigma_A^2}{\sigma_{ph}^2} \times 100$$

Narrow-sense heritability is always less than or equal to broad-sense heritability because narrow-sense heritability includes only additive effects, whereas broad-sense heritability is based on all genetic effects.

The usefulness of broad- vs. narrow-sense heritability depends on the generation and reproductive system of the particular population. In general, narrow-sense heritability is more useful than broad-sense heritability since only additive gene action can normally be transmitted to progeny. This is, because in systems with sexual reproduction, only gametes (alleles) but not genotypes are transmitted to offspring. In contrast, in case of asexual reproduction, genotypes are transmitted to offspring.

Table 6 Comparison of broad- and narrow-sense heritability.

Broad-sense heritability	**Narrow-sense heritability**	
Symbols used	H^2, H, h_b^2, or h_B^2	h^2, h_n^2, or h_N^2
Predictor of Gain	Poor	Better
Genetic Variance	Additive, dominance, and epistatic	Additive only
Generation	Early	Later
Reproductive System	Self-pollinated or cloned population	Cross-pollinated

Estimating Heritability

As the formulas presented above indicate, heritability is calculated from estimates of the components of phenotypic variation: genetic, environmental, and genetic x environmental interactions. Two main approaches are described here to help estimate the contribution of different G, E, and GxE components and for calculating heritability. One approach focuses on eliminating one or more variance component, while the other focuses on comparing the resemblance of parents and offspring. These estimates can be determined from an analysis of variance or regression analysis of the character performance of a population grown in several environments (multiple locations and/or years).

Several Environments

Testing a character's performance in multiple environments (e.g., more than one location and/or years) is essential to get an accurate estimate of the environmental effects on the character. Test environments should be either random or representative of the target environment the type of environment for which the cultivar under development is intended. Heritability is based on variance, the average of the squared deviations from the mean—a statistical measure of how values vary from the mean. Testing in a single environment provides no measure of variance. The greater the number of environments used in the character's evaluation, the better is the reliability of the variance estimate, as well as the heritability estimate. Without adequate testing in multiple environments, heritability estimates may be misleading.

When evaluating different genotypes for a specific character, if the genotypes vary widely in response to differing environments, environmental variance will be relatively high and heritability for the character low. Conversely, if the different genotypes perform in a similar manner across environments, e.g., certain genotypes are always among the best and others always the poorest regardless of environment, environmental variance will be low and heritability high.

Estimation Using Analysis of Variance (ANOVA)

Estimating heritability from an analysis of variance provides a way to measure the relative contributions of two or more sources of variability.

Analytical Tools

Several analytical procedures are commonly used to sort out the sources of variation in the sample, to determine the relationship among factors contributing to the variability, and to estimate the heritability of the character.

Analysis of Variance — this procedure identifies the relative contribution of the sources of observed variation in the sample. Sources of variation may include environment, replication, or genotypic effects. The portion of variation that cannot be attributed to known causes is called "error."

Analysis of Variance table.

Source of Variation	Expected Mean Squares
Genotypes	$\sigma_e^2 + r\sigma_g^2$
Error	σ_e^2

In this example, phenotypic variance is explained by differences in genetic composition, as well as to unknown factors. r stands for the number of replications.

$$\sigma_{ph}^2 = \sigma^2 + r\sigma_g^2$$

The particular steps involved in this procedure and the analysis of variance table that results depend on the design of the experiment. Analysis of variance procedures and interpretation are discussed in the Quantitative Methods course.

Estimation using Parent-Offspring Regression

Heritability can also be estimated by evaluating the similarities between progeny and parent performance using regression analysis. This analysis is based on several assumptions.

- The particular character has diploid, Mendelian inheritance.
- There is no linkage among loci controlling the character of interest, or the population is in **linkage equilibrium**.
- The population is random mated.
- Parents are not inbred.
- There is no environmental correlation between the performance of parents and progeny (to avoid violating this last assumption, randomize parents and progeny within replications; i.e., do not test them in the same plot).

The linear regression model is:

$$Y_i = a + bX_i + e_i$$

where:

- Y_i = phenotypic value of progeny of the i^{th} parent
- a = mean phenotypic value of all parents tested
- b = regression coefficient (slope of the line)
- x_i = phenotypic value of the i^{th} parent
- e_i = experimental error in the measurement of X_i

Analytical Tools

Several analytical procedures are commonly used to sort out the sources of variation in the sample, to determine the relationship among factors contributing to the variability, and to estimate the heritability of the character.

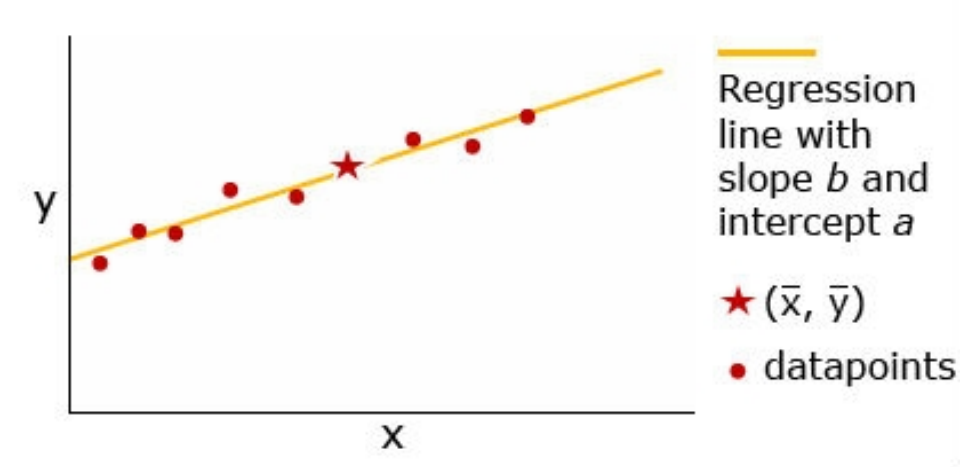
Regression — this procedure examines the strength of the relationship between factors or the

influence one factor has on another. The linear regression procedure fits a straight line to a scatterplot of data points. The general equation for the regression line is:

$$y = a + bx$$

where:

- y = response or dependent variable
- x = predictor or independent variable
- a = y-intercept of the line
- b = regression coefficient, the slope of the line



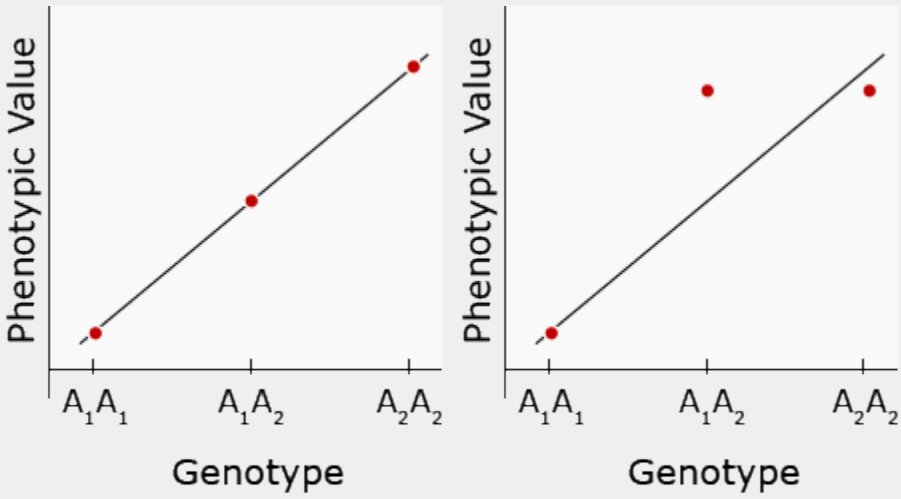
The regression line always passes through the point $(\bar{x}, \bar{y})$. Relative to the total spread of the data, if most of the datapoints lie on or very near the line, there is a strong relationship between the predictor and response variables—x has a strong influence on y. In contrast, the fewer the points that fall on or near the line, the less influence x has on y. A cautionary note: although the x variable may have strong influence on the y variable, x may not be the cause of the y response, nor the sole factor influencing y.

Regression can be used to assess the relative effect of environment on phenotypic value or to obtain information about gene action. The relative scatter about the regression line of a plot of genotype (the predictor variable, x) against phenotypic value (the response variable, y), provides information about gene action. For example, regression analysis of the following two examples suggests that the gene action in example 1 (left panel) is additive (no dominance), whereas there is a complete dominance (by the A_2 allele) in the case of example 2 (right panel) (Fehr, 1987).

When the heterozygous genotype has a value midway between the two homozygotes and thus all

three genotypic values fall on the linear regression line the only gene action contributing to the phenotype is additive.

Regression analysis of phenotypic values of progeny (y) against parents (x) provides useful information about the degree of similarity of progeny to the parents.



Regression analysis may indicate the type of gene action.

Alternative Formula

An alternative formula for calculating the regression coefficient, b, is

$$b = \frac{\Sigma(X - \bar{X})(Y - \bar{Y})}{\Sigma(X - \bar{X})^2}$$

where:

- b = regression coefficient
- X = parent values
- Y = progeny values

The performance of the progeny is a function of the genetic factors inherited from the parents. (Assume that "parent" means either a random plant or line from a population.) Thus, X,

the parent value, is the independent variable, and Y, the progeny value, is the response or dependent variable.

Regression Coefficient

What does the regression coefficient, b, tell us?

If b = 1, then

- gene action is completely additive,
- negligible environmental effect,
- and negligible experimental error.

The smaller the value of b, the less closely the progeny resemble their parent(s), indicating

- greater environmental influence on the character,
- greater dominance and/or epistatic effects, and/or
- greater experimental error.

An analysis of variance will provide estimates of the relative influence of genetics, environment, and experimental error.

The type of heritability and the specific formula used to estimate it depends on the type of progeny evaluated.

Types of Progeny

The population's reproductive mode and mating design determine the type of heritability and the formula used to calculate the estimate.

Selfed progeny

1. F_2 plants are self-pollinated to obtain the F_3. All the alleles in the F_3 come from the F_2 parent. Evaluate the character performance of the F_3
2. Regress the performance of the F_3 on the performance of their F_2 parents.
3. The regression coefficient, b, is equal to H^2, broad-sense heritability because its genetic

variance includes dominance, additive, and epistatic effects.

$H^2 = b \times 100$

4. Since it is difficult to obtain information about gene action (dominance, epistasis, additive) in self-pollinated populations, narrow-sense heritability is a poor predictor of genetic gain and rarely used in these populations. Inbreeding causes an upward bias in the heritability estimate.

Full-sib progeny

1. Two random F_2 plants are mated. Half of the alleles in the F3 come from one parent and half from the other. Evaluate the character performance of the F3.
2. Determine the **mid-parent value** of the two parents. Mid-parent value, X = $(x_1 + x_2)$ /2
3. Regress the progeny on the mid-parent value. The regression of progeny on the mid-parent value is

$$\frac{\frac{1}{2}V_A}{\frac{1}{2}V_P} = \frac{\frac{1}{2}\sigma^2_A}{\frac{1}{2}\sigma_{ph}} = \frac{\sigma^2_A}{\sigma^2_{ph}}$$

Since the progeny have both parents in common, only additive variance is included, so the regression coefficient, b, is equal to narrow-sense heritability.

$$h^2 = b \times 100$$

Half-sib progeny

1. Open-pollinate F_2 plants. Seed will be harvested from each F_2 individual separately. Progenies from the same F_2 plant have the maternal parent (the respective F_2 plant) in common, while the paternal parent is pollen from the whole F_2 population. Thus all offspring from an F_2 plant after open pollination are “half-sibs”.
2. Regress the performance of the half-sib progenies on their parents.

$$b = \frac{\sigma^{xy}}{\sigma^2_x}$$

where:

σ^{xy} = covariance between parents, x, and their progeny, y

σ^2_x = phenotypic variation among parents

3. The covariance between parents and progeny includes additive variance and some forms

of additive epistasis (usually negligible), but no dominance variance. Thus, narrow-sense heritability can be estimated. The regression coefficient, b, is equal to half the heritability value.

4. Multiply the regression coefficient by 2 to obtain narrow-sense heritability.

$$h^2 = 2b \times 100$$

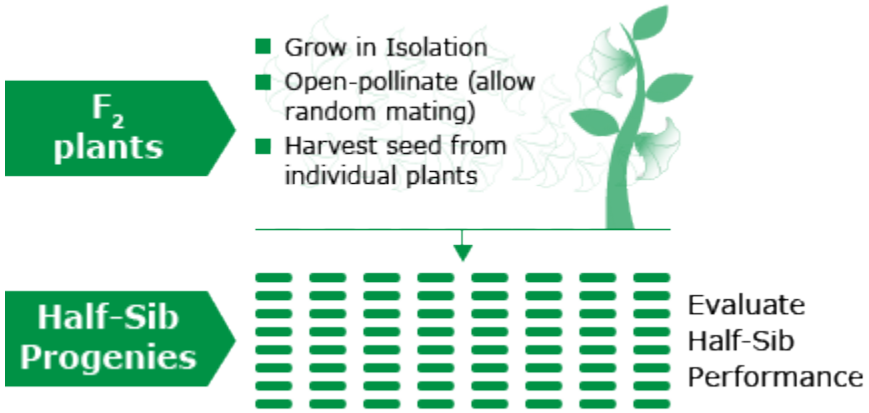


Fig. 12 Half-sib progeny.

Heritability Influences

Heritability is not an intrinsic property of a trait or a population. As we've seen, it is influenced by:

- population — generation, reproductive system, and mating design
- environment — locations and/or years
- experimental design — experimental unit (plant, plot, etc.), replication, cultural practices, techniques for data gathering.

Heritability can be manipulated by increasing the number of replicates and number of environments sampled (in space and/or time). Genetic variance can be increased by using diverse parents and by increasing the selection intensity. Heritability is only an indicator to guide the breeder in making selections and is not a substitute for other considerations, such as breeding objectives and resource availability.

Genetic Advance from Selection

Evolution

Estimating Response to Selection

Evolution can be defined as genetic change in one or more inherited trait that takes place over time within a population or group of organisms. Plant breeders can use quantitative genetics to predict the rate and magnitude of genetic change. The amount and type of genetic variation affects how fast evolution can occur if selection is imposed on a phenotype.

The amount that a phenotype changes in one generation is called the selection response, **R**. The selection response is dependent on two factors—the narrow-sense heritability and the selection differential, **S**. The selection differential is a measure of the average superiority of individuals selected to be parents of the next generation.

$$R = h^2 S$$

The above equation is often called the "breeder's equation". It shows the key point that **response to selection** increases when either the **heritability** of the trait or the **strength of the selection** increases.

Realized Heritability

In an experiment, the observed response to selection allows the calculation of an estimate of the narrow-sense heritability, often called the realized heritability. A low h^2 (<0.01) occurs when offspring of the selected parents differ very little from the original population, even though there may be a large difference between the population as a whole and the selected parents. Conversely, a high h^2 (> 0.6) occurs when progeny of the selected parents differ from the original population almost as much as the selected parents.

In the figure in the next screen, it can be seen that the selection differential (S) in each generation is the difference between the mean of the entire original population and the mean of group of individuals selected to form the next generation. In contrast, the response to selection (R) indicates the differences in population means across generations. The value of R is the difference between the mean of the offspring from the selected parents and the mean of the entire original population:

$$S = \bar{T}_S - \bar{T}$$

$$R = \bar{T}_O - \bar{T}$$

where:

- T = mean of the entire original population
- T_S = mean of selected parents
- T_O = mean of the offspring of the selected parents

Adaptive Value

The proportionate contribution of offspring of an individual to the next generation is referred to as fitness of the individual. Fitness is also sometimes called the adaptive value or selective value. Note in the figure that the non-selected members of the population do not contribute to the next generation and that selection over time reduced the variance of the population.

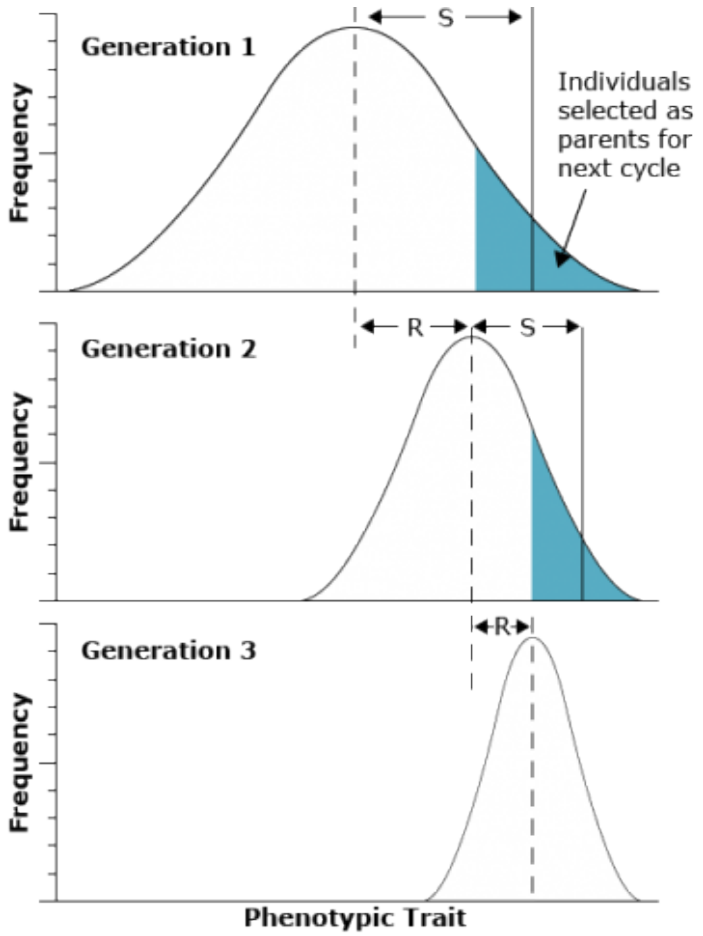


Fig. 13 How population means shift as a result of selection.

Types of Selection

Artificial selection refers to selective breeding of plants and animals by humans to produce populations with more desirable traits. Artificial selection is typically directional selection because it is applied to individuals at one extreme of the range of variation for the phenotype selected. This type of selection process is also called truncation selection because there is a threshold phenotypic value above which the individuals contribute and below which they do not. In contrast, under natural selection in non-managed populations, other types of selection may occur.

Three main types of selection are generally recognized. All three operate under natural selection in natural populations, whereas under artificial selection via selective breeding by humans only directional selection is common.

Directional Selection

Directional selection acts on one extreme of the range of variation for a particular characteristic.

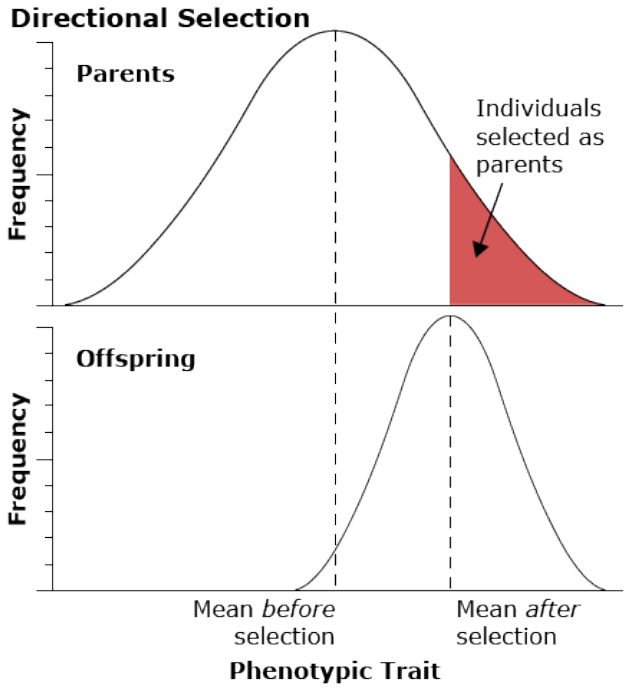


Fig. 14 Directional selection

Stabilizing selection

Stabilizing selection works against the extremes in the distribution of the phenotype in the population. An example of this type of selection is human birth weight. Infants of intermediate weight have a much higher survival rate than infants who are either too large or too small.

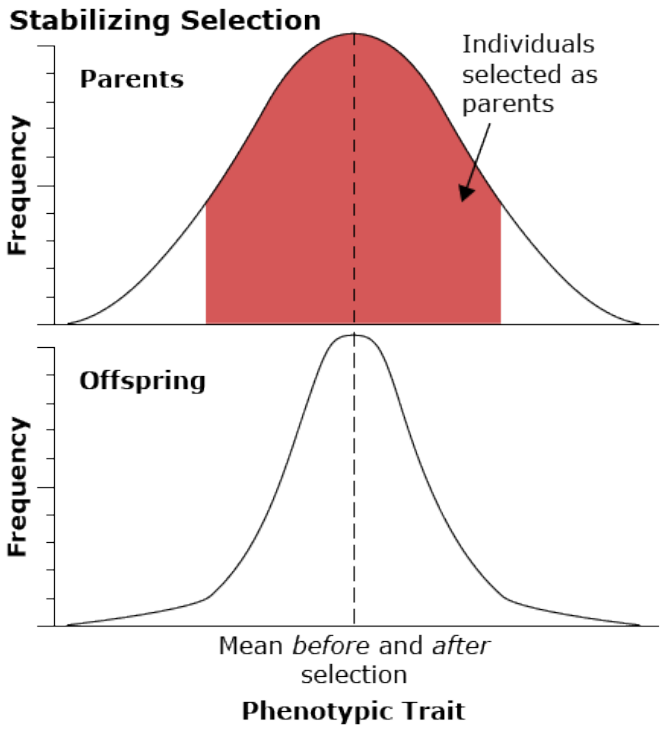


Fig. 15 Stabilizing selection.

Disruptive selection

Disruptive selection favors the extremes and disfavors the middle of the range of the phenotype in the population.

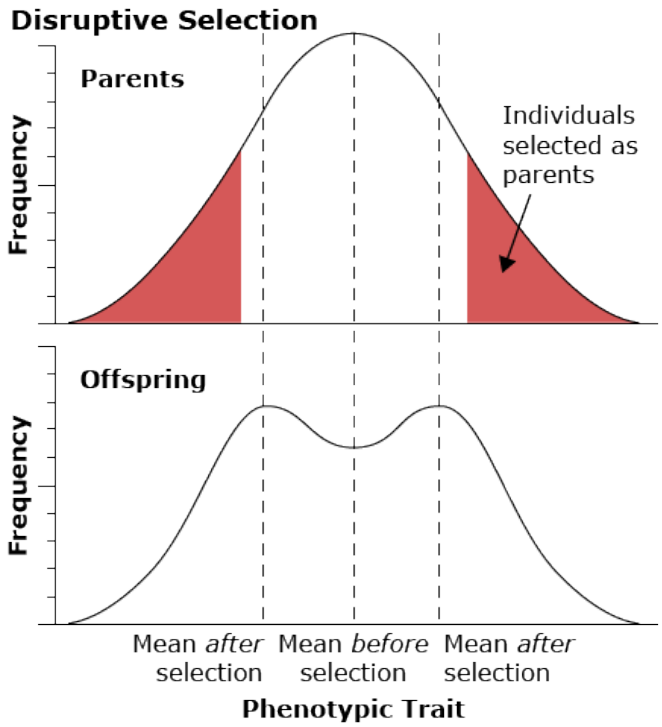


Fig. 16 Disruptive selection.

Genetic Advance from Selection

One of the most famous longest-term selection experiments is a study conducted by University of Illinois geneticists who have been selecting maize continuously for over 100 generations since 1896. They have been changing oil and protein content in separate experiments, selecting for either high or low content. In some cases after multiple generations, they have shifted selection from high to low or vice versa.

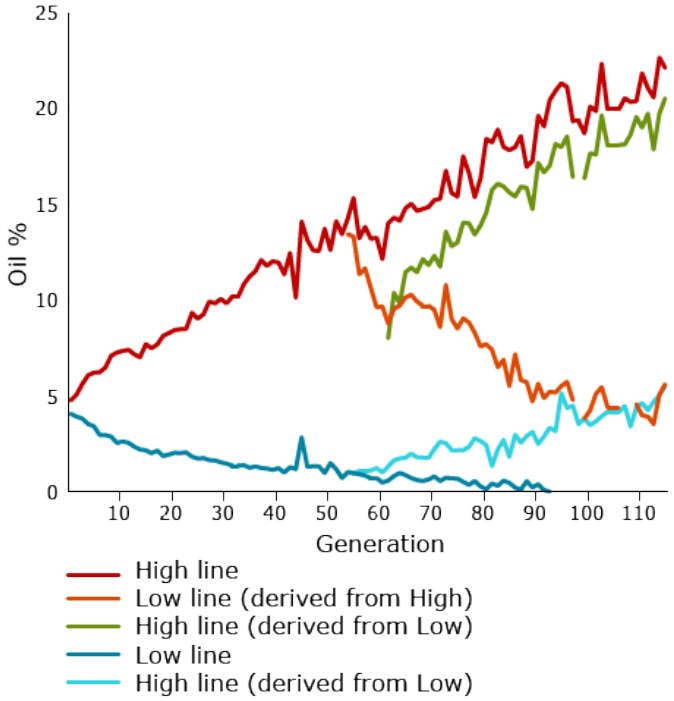


Fig. 17 Responses to 100 generations for high and low oil concentration in maize kernels. Data from Department of Crop Science, University of Illinois at Urbana-Champaign, 2007.

Expected Gain From Selection

Because resources are limited, the breeder's objective is to carry forward as few plants or lines as possible without omitting desirable ones. How does the breeder decide how many and which plants or lines within a population to carry forward to the next generation? The breeder can use heritability estimates to predict the probability that selecting a given percentage of the population or selection intensity, **i**, will result in progress. The expected progress or gain can be calculated using this formula:

$$G_C = (k)(\sqrt{V_P})(h^2) = k\sigma_{ph}h^2$$

where:

- G_c = expected gain or predicted genetic advance from selection per cycle
- k = selection intensity — a constant based on the percent selected and obtained from statistical tables (note that some people use hte i symbol instead of k for selection intensity

- $\sqrt{V_p}$ or σ_{ph} = square root of phenotypic variance (equivalent to standard deviation)
- h^2 = narrow – sense heritability in decimal form (narrow – sense is used for sexually reproduced populations whenever possible, and broad sense heritability, H^2, is used for self – pollinating and asexually reproduce populations)

Caution: The phenotypic values must exhibit a normal, or bell-curve, distribution for G_c to be valid

Selection Intensity Table

As long as the distribution of phenotypic values is normally distributed, selection intensity values (symbolized by k or sometimes i) can be found in statistical tables. The intensity of selection practiced by plant or animal breeders depends just on the proportion of the population in the selected group.

The selection intensity is a standardized selection differential and is a measure of the superiority of the individuals selected as parents for breeding relative to the population from which they were selected. Representative values of k are shown in the table.

Representative selection intensity (k) values.

%	k
1	2.67
2	2.42
5	2.06
10	1.76
20	1.40
50	0.80
90	0.20
100	0

For Your Information

Performance

As will be explained in the next section, a key statistic used to describe populations is the mean performance of the population of genotypes. The mean performance of a population can be described by a combination of values for performance of both homozygous and heterozygous genotypes, as well as the relative frequency of alleles.

For illustration using a locus with two alleles, A_1 and A_2, the genotypic value of the homozygotes is designated as A_1A_1 =+a and A_2A_2 = -a, while the heterozygous genotype is A_1A_2 = d.

The value of a is the performance of a homozygous genotype minus the average performance of the two homozygous genotypes.

+a = A_1A_1 - \frac{(A_1A_1 + A_2A_2)}{2}

$$-a = A_2A_2 - \frac{(A_1A_1 = A_2A_2)}{2}$$

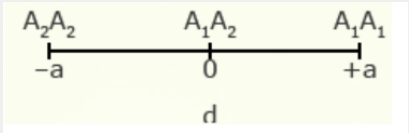
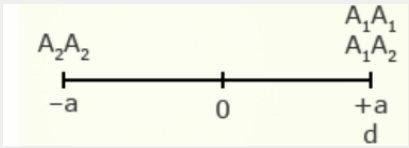
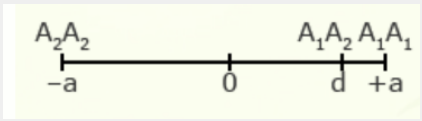
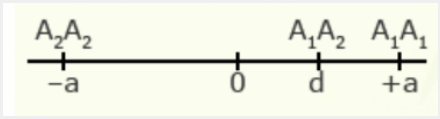
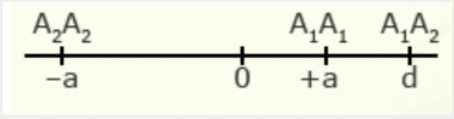
The value of d measures the degree of dominance between alleles, and is the difference between the value of the heterozygote and the mean of the homozygotes.

$$d = A_1A_2 - \frac{(A_1A_1 + A_2A_2)}{2}$$

Degrees of Dominance

Examples of relative genotypic values are given here in depictions showing a and d under different types of gene action.

Gene action depicted using a and d in relation to genotypic values. Adapted from Conner and Harti, 2004.

Gene Action	Degree of Dominance	Relative Genotypic Values
Additive	0	
Complete dominance	1	
Partial dominance	3/4	
Partial dominance	1/2	
Overdominance	2	

The difference between the depictions of partial dominance show an example of how the effect of an allele can vary depending on whether the locus is a major gene or minor. Remember, additivity of genetic effects for quantitative traits does not mean that there are equal effects of all alleles at a locus or all loci affecting the trait.

The study of quantitative traits is sometimes referred to as “statistical genetics” because of its

reliance on statistical methods. In order to understand the inheritance of quantitative characters and the methods applied to these characters, it is essential that you become familiar with fundamental statistics. A basic review is provided here.

When using symbols to represent these population parameters, it is important to distinguish between information about the population and that concerning a sample representing the population.

Genotypic and Breeding Values

Populations can be characterized by the amount and type of genetic variability contained within them. Genetic improvement of a quantitative character is based on effective selection among individuals that differ in what is known as the genotypic value. Variation among the genotypic values represents the genotypic variance of a population.

The genotypic value is the phenotype exhibited by a given genotype averaged across environments. A related concept is the breeding value, which is the portion of the genotypic value that determines the performance of the offspring. **Genotypic value** is property of the genotype and therefore is a concept that describes the value of genes to the individual, whereas **breeding value** describes the value of genes to progeny and therefore helps us understand how a trait is inherited and transmitted from parents to offspring . Remember that only additive genetic effects can be passed on to progeny. Non-additive genetic effects and environmental effects cannot be inherited by offspring.

References

Barbour, M.G, J.H. Burk, F.S. Gilliam, W.D. Pitts and M.W. Schwartz. 1999. Terrestrial Plant Ecology. 3rd edition. Benjamin Cummings, San Francisco, CA.

Clausen, J., D.D. Keck, and W. Hiesey. 1940. Experimental studies on the nature of species. I. Effects of varied environments on western North American plants. Carnegie Inst. Wash. Publ. 520.

Clausen, J., D. D. Keck, and W. M. Hiesey. 1948. Experimental studies on the nature of species. III. environmental responses of climactic races of Achillea. Carnegie Inst. Wash. Publi. 581.

Conner, J.K. and D.L. Hartl. 2004. A Primer of Ecological Genetics. Sinauer Associates, Sunderland, MA.

Falconer, D.S., and T.F.C. Mackay. 1996. Introduction to Quantitative Genetics. 4th edition. Longman Publ. Group, San Francisco, CA.

Hallauer, A. R. and J. B. Miranda. 1988. Quantitative Genetics in Maize Breeding. 2nd Edition. Iowa State University Press, Ames, IA.

Hill, W.G. 2005. A century of corn selection. Science 307: 683-684.

Pierce, B. A. 2008. Genetics: A Conceptual Approach. 3rd edition. W.H. Freeman, New York.

Rausher, M.D.. 2005. Example of Clausen, Keck, and Hiesey Experiment, Lecture 1-Lineages, Populations, and Genetic Variation. Online lecture notes from course on Principles of Evolution.

Department of Crop Science, University of Illinois at Urbana-Champaign. Values obtained for protein in the strains selected for oil and the values for oil obtained for the strains selected for protein each generation (1896-2004). 2007.

How to cite this chapter: Beavis, W., L. Merrick, K. Meade, A. Campbell, D. Muenchrath, and S. Fei. (2023). Inheritance of quantitative traits. In W. P. Suza, & K. R. Lamkey (Eds.), *Crop Genetics*. Iowa State University Digital Press. DOI: 10.31274/isudp.2023.130

Mutations and Variation

Laura Merrick; Arden Campbell; Deborah Muenchrath; Shui-Zhang Fei; and William Beavis

Introduction

Mutations are the ultimate source of all genetic variation. Mutations can occur at all levels of genetic organization, classified mainly as either **chromosome mutations** or **genome mutations**. Chromosome mutations are discussed in this module. Chromosome alterations involve either single **nucleotides** or fragments of chromosomes and are either small-scale (one or a few nucleotides substituted, inserted, or deleted) or large-scale (deletions, insertions, inversions, or translocations involving large segments of chromosomes or duplications of entire genes). Genome mutations—involving changes in number of whole chromosomes or sets of chromosomes—will be covered separately in the module on Ploidy—Polyploidy, Aneuploidy, Haploidy.

Genetic variation—dissimilarity between individuals attributable to differences in genotype—that is generated by mutations is acted upon by various evolutionary forces. Evolutionary processes that alter species and populations include selection, gene flow (migration), and genetic drift—whether or not plants are cultivated or wild. **Evolution** can be defined as a change in gene frequency over time. The way that plants evolve is dependent on both genetic characteristics and the environment they face.

Genetic variation results from differences in DNA sequences and, within a population, occurs when there is more than one allele present at a given locus. Major processes that affect heritable variation in crop plants are topics emphasized throughout the lessons of this course. Changes in gene frequencies within populations caused by natural selection can lead to enhanced adaptation, while changes caused by human-directed selection can facilitate the development of useful genetic variability and selection of superior genotypes. Selection is the differential reproduction of the products of recombination—both within and between chromosomes.

Genetic Resources

Historically plant breeders seeking sources of variability were constrained in choice of parental materials or **plant genetic resources** that were interfertile within closely related **gene pools.** But a range of new techniques such as mutagenesis, **genetic engineering** (**transgenic** or transformed plants), and *in vitro* methods (tissue culture, doubled haploids, induced polyploids) expand the source and scope of variability that can be used in crop improvement.

Our expanding understanding of the molecular basis of genetics has provided insights and technologies that further not only our basic understanding of genes and their regulation, but also provide additional tools for crop improvement. Molecular techniques enable breeders to generate genetic variability, transfer genes between unrelated species, move synthetic genes into crops, and make selections at the molecular, cellular, or tissue levels. Combining these laboratory techniques with conventional field approaches can shorten the time required to develop new or improved cultivars. The importance and application of molecular technologies are rapidly increasing.

These topics mentioned above—mutations, gene expression, genetic markers, sources of genetic variation, genetic engineering, and molecular breeding methods—will be briefly mentioned in this module, but covered in greater detail in the later courses including Plant Breeding Methods, Molecular Genetics, and Biotechnology and Molecular Plant Breeding.

Learning Objectives

- Recognize how mutations are classified and inherited, as well as how mutations affect structure, processes, and products of genes and chromosomes.
- Understand the basic principles of transcription and translation.
- Become familiar with sources of genetic variation for cultivated plants, including crop gene pools and genetic engineering methods.

Mutations as Heritable Change

Without heritable variation, any trait favored by selection will not be passed on to offspring. Mutation is defined as heritable change in genetic information. Mutations entail modification of the nucleotide sequence of DNA and consist of any permanent alteration of a DNA molecule that can be passed on to offspring. DNA is a highly stable molecule and it replicates with a high degree of accuracy. However changes in DNA structure and replication errors can occur. Mutation involves modifications in the sequence of bases in DNA transmitted through mitosis and meiosis.

Nucleotides

A **nucleotide** consists of a sugar molecule (**ribose** in RNA or **deoxyribose** in DNA) attached to phosphate group and a nitrogen-containing base. In DNA or RNA molecules, each strand has a backbone of sugar and phosphate groups (Fig. 17).

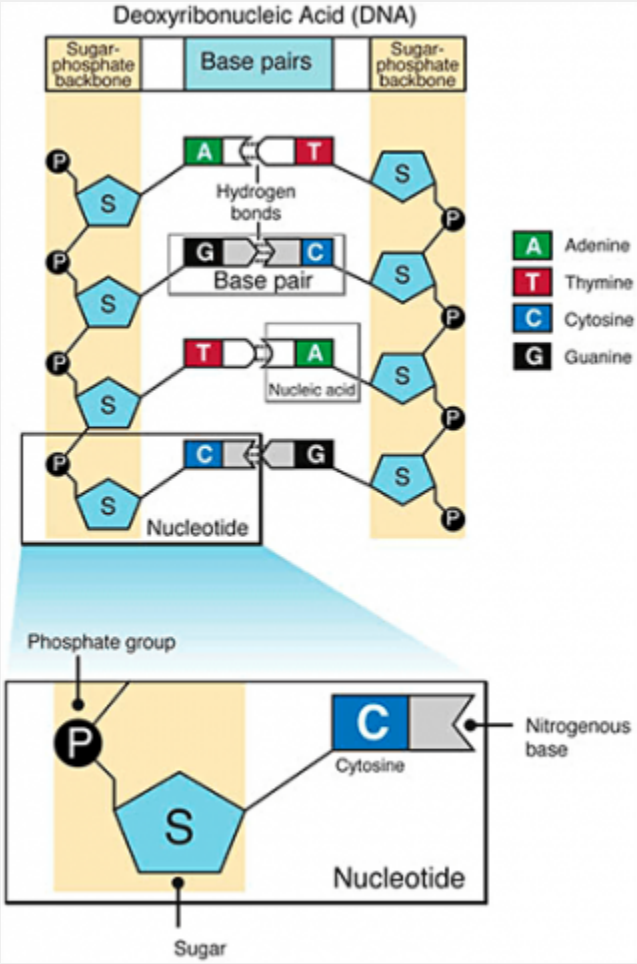


Fig. 12 Nucleotides are the basic building block of the nucleic acids such as DNA. Illustration from NIH-NHGRI, 2011.

CHEMICAL BASES IN DNA AND RNA

Two of the four nitrogenous bases in DNA—**adenine** (Fig. 13) and **guanine** (Fig. 14) are known

as **purines** and the other two—**cytosine** (Fig. 15) and **thymine** (Fig. 16) are **pyrimidines**. Adenine, guanine, and cytosine are also found in RNA. Another pyrimidine known as **uracil** (Fig. 17) is the base used in RNA in place of thymine.

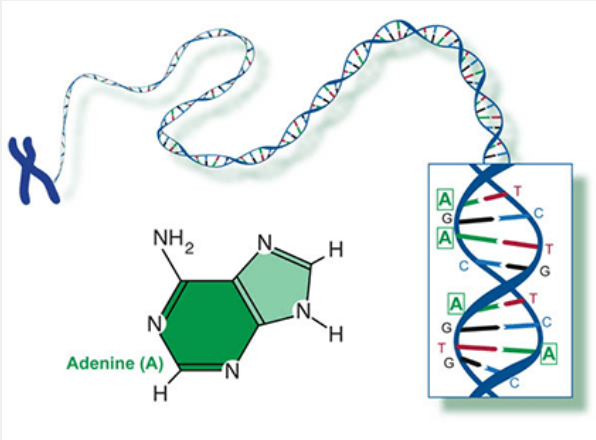


Fig. 13 Adenine, a nitrogenous base in DNA and RNA, is a type of purine. Illustration from NIH-NHGRI, 2011.

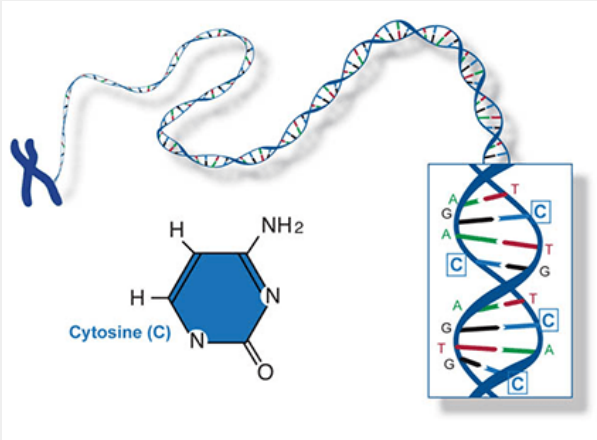


Fig. 15 Cytosine, a nitrogenous base in DNA and RNA, is a type of pyrimidine. Illustration from NIH-NHGRI, 2011.

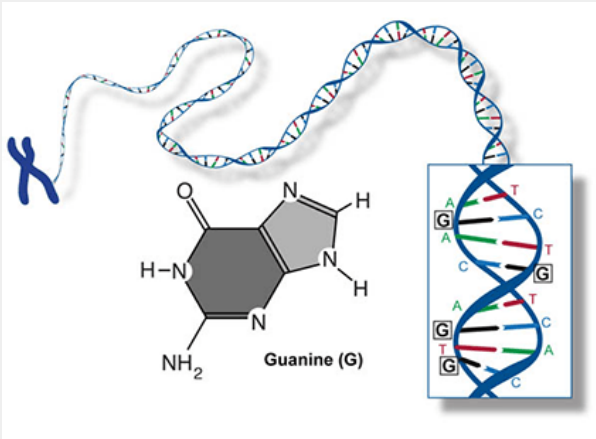


Fig. 14 Guanine, a nitrogenous base in DNA and RNA, is a type of purine. Illustration from NIH-NHGRI, 2011.

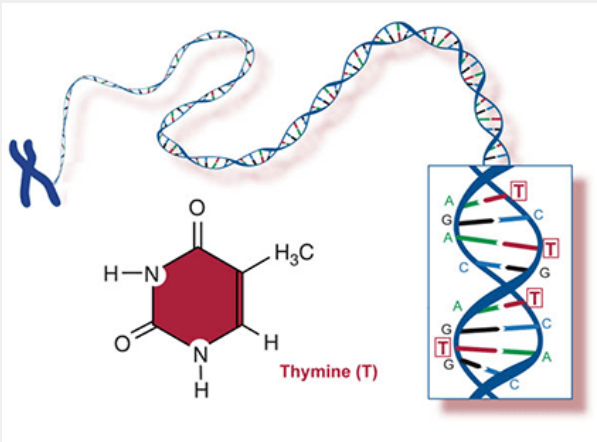


Fig. 16 Thymine, a nitrogenous base in DNA, is a type of pyrimidine. Illustration from NIH-NHGRI, 2011.

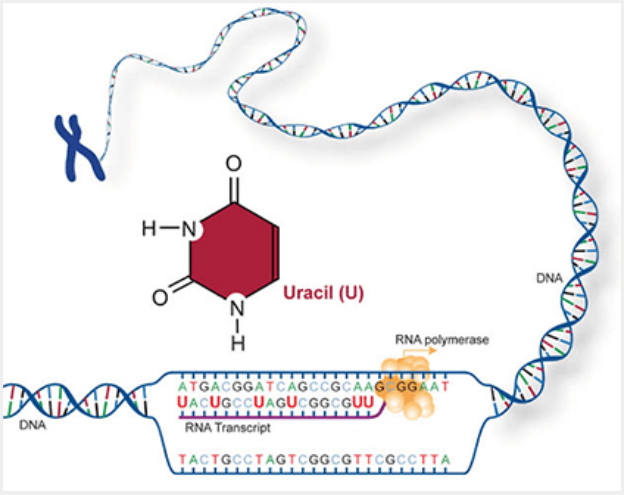


Fig. 17 Uracil, a nitrogenous base in RNA, is a type of pyrimidine. Illustration from NIH-NHGRI, 2011.

Some mutations occur in loci that encode for gene products such as proteins, and thus they may affect the processes of **transcription**, **translation**, or **gene expression**—processes that happen during the creation of proteins from the genetic code in DNA. But mutations also can occur in parts of the genome that do not code for any gene products (called noncoding DNA) or sequences that serve to control regulatory functions in the cell or chromosomes. For most loci, mutation changes allelic frequencies at a very slow rate and therefore consequences are negligible. Mutations may or may not change the phenotype of an organism. The majority of mutations that do occur are neutral in their effect and therefore do not have an influence on fitness. Some mutations are beneficial. But mutations can have deleterious effects, causing disorders or death.

Amino Acids and Proteins

Amino acids are a set of 20 different molecules used to build proteins. A **peptide** is one or more amino acids linked by chemical bonds (termed peptide bonds). Linked amino acids form chains of **polypeptides** (Fig. 18). The amino acid sequences of proteins are encoded in genes.

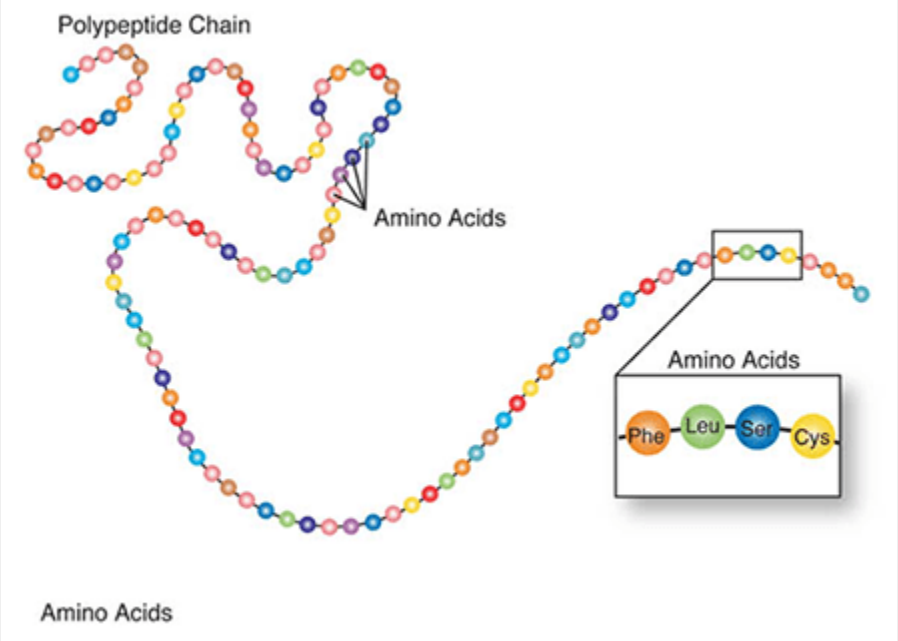


Fig. 18 Amino acids bond together with peptide bonds to form long chains of polypeptides. Illustration from NIH-NHGRI, 2011.

One or more polypeptides form the building blocks of proteins (Fig. 19). Proteins perform a variety of roles in cells.

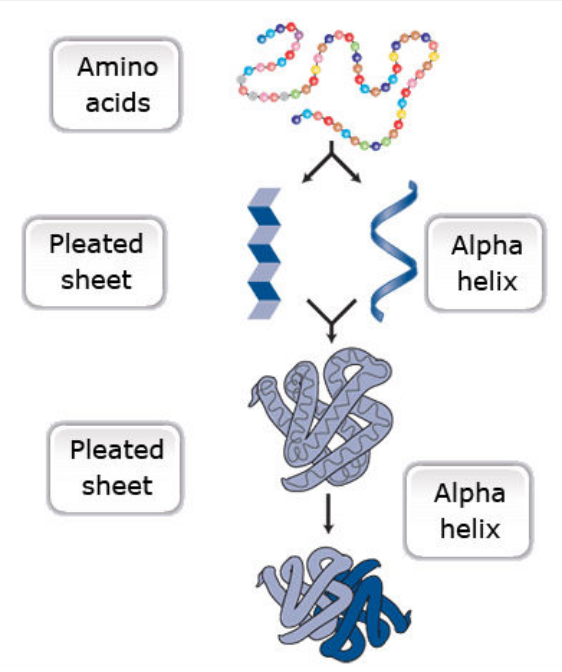


Fig. 19 Polypeptides are combined and arranged to form proteins. Illustration from NIH-NHGRI, 2011.

Primary protein structure is a sequence of a chain of amino acids.

Secondary protein structure occurs when the sequence of amino acids is linked by hydrogen bonds.

Tertiary protein structure occurs when certain attractions are present between alpha helices and pleated sheets.

Quaternary protein structure is a protein consisting of more than one amino acid chain

Types of Mutations

Classification of Mutations

Mutations can occur at all levels of genetic organization, ranging from simple base nucleotide pair alterations to shifts and rearrangements in sequences of nucleotides along fragments of chromosomes to changes in the number and structure of whole chromosomes.

A mutation is a change from one hereditary state to another, e.g., allele ***A*** mutates to allele ***a***. For a given locus, the normal allele is referred to as the '**wildtype**'. Mutations are usually recessive and therefore their effects are hidden in heterozygotes. There are a number of common ways to classify mutations, including the following:

- causal agent
- rate or frequency of occurrence
- kind of tissue involved and its type of inheritance
- impact on fitness or function, or
- molecular structure and scale of the mutation.

Spontaneous vs. Induced Mutations

Depending on the cause, mutations can be either spontaneous or induced:

- **Spontaneous mutations** occur naturally with no intentional exposure to a **mutagen**. Spontaneous mutations can result from copying errors made during cell division.
- **Induced mutations** are caused by mutagens, either chemicals or radiation.

RARE VS. RECURRENT MUTATIONS

Recall from the module on Population Genetics, mutational events in a population can be classified into two categories based on frequency of occurrence:

Rare mutations (also called non-recurrent mutations) are defined as those that occur infrequently in populations. Rare mutations are usually recessive and occur in a heterozygous condition so that their effect on the phenotype is not apparent. Rare mutations will usually be lost from populations due to random genetic drift.

Recurrent mutations are defined as those that occur repeatedly and thus can possibly cause a change in gene frequency in populations. For a given locus, the rate of allele **A** mutating to allele **a** can be given as the frequency u per generation; **a** mutates to **A** at a rate v.

With the frequency of **A** symbolized as p and that of **a** symbolized as q, then at equilibrium, $pu = qv$, or $q = p/(u + v)$ (see the Equation below).

$$pu = qv$$

$$q = \frac{u}{v + u}$$

Mutation Rates

Falconer and Mackay (1996) summarize the following key points about mutation rates and their frequency in populations:

- normal **spontaneous mutations** alone can produce only very slow changes of allele frequency;
- **mutation rates** are generally quite low for most loci in most organisms, occurring about 10^{-5} to 10^{-6} per generation or, stated another way, about 1 in 100,000 to 1 in 1,000,000 gametes carry a newly mutated allele at any locus;
- with respect to equilibrium in both directions (*u* and *v*) in natural populations, **forward mutation** (from wildtype to mutant; *u*) is much more frequent than **reverse mutation** (from mutant to wildtype; *v*); and
- an equilibrium state known as the **mutation-selection balance** can maintain deleterious alleles at low frequency; selection acts to eliminate deleterious recessive alleles, but very slowly when the allele frequency is low; even if the elimination process of selection is slow, an equilibrium occurs if mutation creates new copies of the deleterious allele.

Somatic vs. Germinal Mutations

Plants are multicellular organisms, but mutation typically starts from a single cell. There are two broad categories of mutations that are classified according to the type of cell tissue involved.

Somatic mutations occur in somatic tissue, which does not produce gametes. Somatic cells divide by mitosis and therefore through that process, mutations can be passed on to daughter cells. Somatic mutations may have no effect on the phenotype if their function is covered by that of normal cells. However somatic cells that stimulate rapid cell division are the basis for tumors in plants and animals. Somatic mutations usually occur as single events (typically in a single cell) in multicellular organisms or organs that lead to **chimera**, which is a part of a plant with a genetically different constitution as compared to other parts of the same plant. Somatic mutations are not transmitted to progeny (Fig. 1).

Germinal or **germ-line mutations** occur in reproductive cells that produce gametes, and

therefore can be passed on to future generations. Germ cells or gametes are formed by meiosis. If a germinal mutation is inherited, then it can be carried in all of the somatic and germ-line cells of the offspring (Fig. 1).

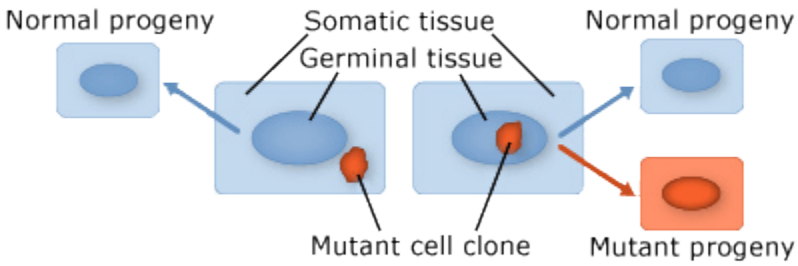


Fig. 1 Somatic vs. germinal mutations.

Effects of Mutation on Fitness or Function

Mutations can affect fitness in various ways and can therefore be classified based on their effect on individual fitness:

- **Deleterious mutations** are those that are harmful and have a negative effect on phenotype, decreasing the fitness of the individual.
- **Advantageous mutations** are those that are beneficial and have a positive or desirable effect on phenotype, increasing the fitness of the individual.
- **Neutral mutations** have neither beneficial nor harmful effects.
- **Lethal mutationsare detrimental and lead to the death of the organism** when present.

Mutations can also be classified by their effect on gene function:

- **Loss-of-function mutations** either result in a gene product that has less function or one that has no function. Phenotypes associated with loss-of-function mutations are usually recessive. Many of the mutations that are associated with crop domestication from wild progenitors involve loss-of-function alleles (Gepts 2002).
- **Gain-of-function mutations** result in a gene product that has novel function. Altered phenotypes associated with gain-of-function mutations are usually dominant. Many of the changes in crop plants brought about by genetic engineering involve gain-of-function mutations (Gepts 2002).

However, it is important to underscore that not all mutations occur in genes or protein-coding regions of the chromosome, nor do all mutations that do occur in genes lead to altered proteins.

Point vs. Chromosomal Mutations

Mutations are often divided into those that affect a single gene, termed a **gene mutation**—also sometimes called a point mutation—and those that affect the structure of chromosomes, called a **chromosomal mutation**. These latter two classes of mutations will be covered in more detail after the concept of gene expression is introduced in the following section.

Point Mutation

Point or Gene Mutation

A point mutation is when a single base pair (or just a few) is altered (Fig. 2), an alteration at a "micro" level. There are two general types of point mutations: **substitutions** or **insertions** and **deletions** (the latter two are collectively called INDELs).

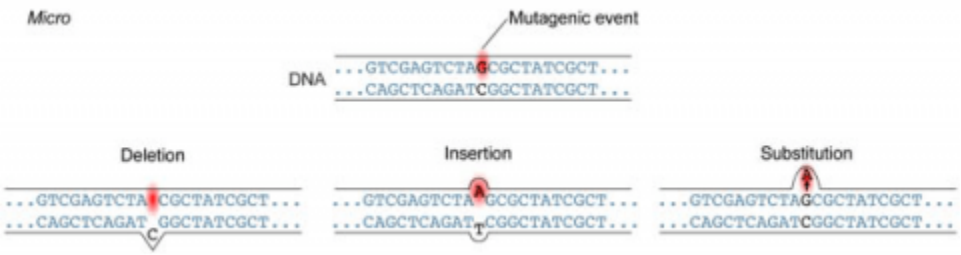
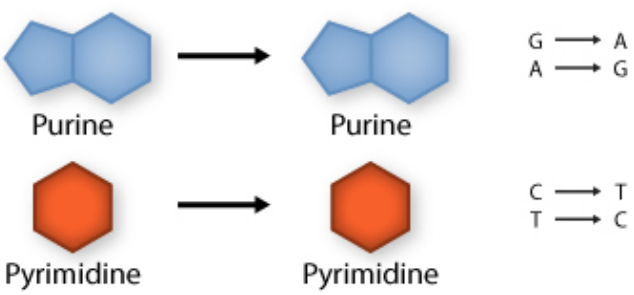
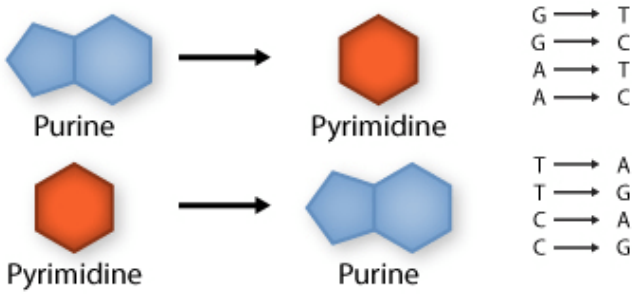


Fig. 2 Point mutations involve base pair deletions, insertions, or substitutions. Illustration adapted from NIH-NHGRI, 2011.

Base pair substitutions involve an alteration of a single nucleotide in the DNA. A substitution mutation can entail either a **transition** or a **transversion**:



Transition: purine replaced by a different purine or pyrimidine by a different pyrimidine



Transversion: purine replaced by a pyrimidine or pyrimidine by a purine

Missense Mutations

Some substitution mutations have no effect on the protein coded for. One reason is because of the redundancy of the genetic code (recall that about one fourth of all base pair substitutions code for the same amino acid; such mutations are termed **silent mutations** since there is no change in amino acid that results from the substitution). Another reason for lack of effect is that even if a change in amino acid occurs (termed **missense mutation**), it may have no actual influence on the function of a protein (Fig. 3). Also any mutation located within a non-coding region of the chromosome will not be translated into a protein. Lastly, an altered gene may be masked by other normal copies of the gene present in the genome.

In certain cases, point mutations can have a significant effect—particularly when a substitution produces a stop codon so that the alteration causes the protein synthesis to halt before the protein is entirely translated, altering the entire structure. These are called **nonsense mutations** (Fig. 3).

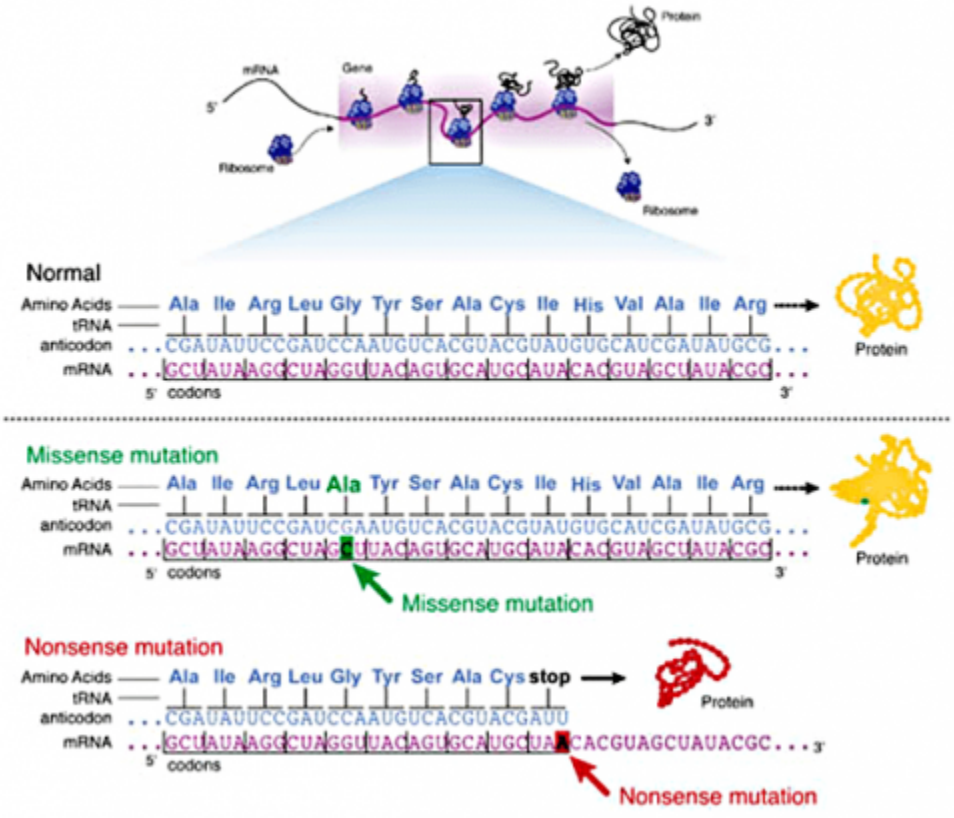


Fig. 3 Missense mutations may or may not affect protein synthesis, whereas nonsense mutations typically have a significant effect. Illustration from NIH-NHGRI, 2011.

Frameshift Mutation

Base pair insertions and deletions are additions (INDELs) or losses of one to several nucleotide pairs in a gene (Fig. 2). Mutations that are insertions and deletions tend to have a much greater effect than do mutations that are base pair substitutions because they disrupt the normal reading frame of trinucleotides. Recall that each group of three bases corresponds to one of 20 different amino acids used to build a protein. Mutations involving base pair insertions and deletions are often therefore referred to as **frameshift mutations**. Under these circumstances the DNA sequence following the mutation is read incorrectly (Fig. 4).

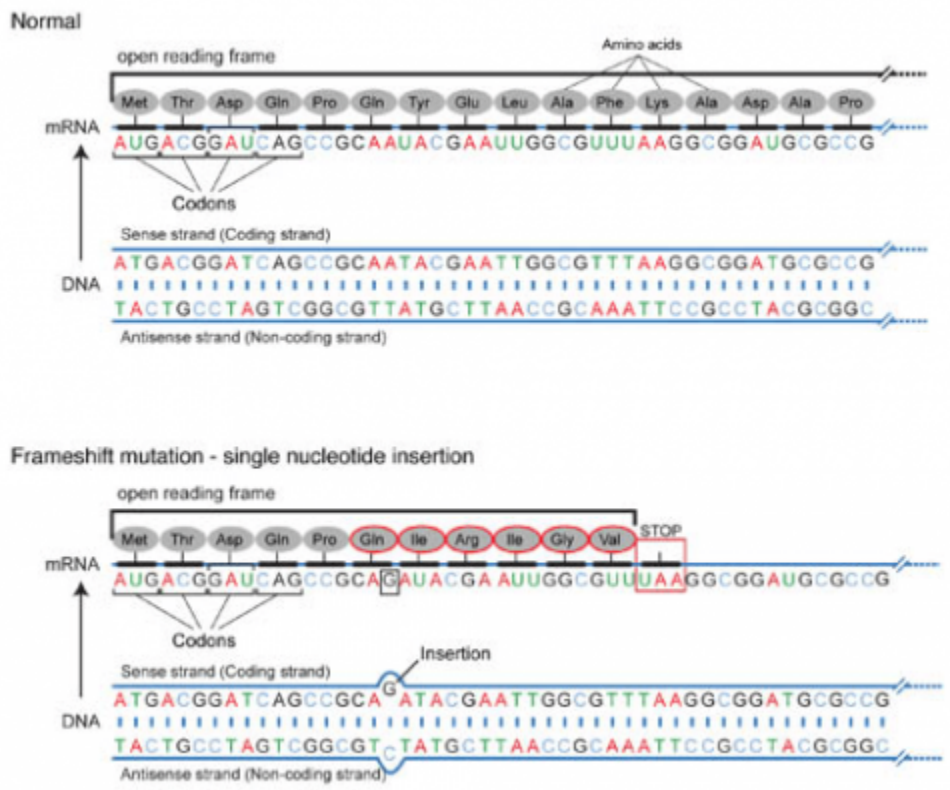


Fig. 4 Frameshift mutation. Illustration from NIH-NHGRI, 2011.

Chromosomal Mutation

MUTATIONS INVOLVING CHROMOSOME SEGMENTS

Different cells of the same organism and different individuals of the same species generally have the same number of chromosomes, and homologous chromosomes are typically uniform in number and in the arrangement of genes along them. However, mutations can occur that alter the number or structure of chromosomes.

Changes involving chromosomal rearrangements entail the following basic types: **deletions**, **duplications**, **insertions**, **inversions**, **substitutions**, and **translocations**—alterations that occur at a "macro" level (Fig. 5).

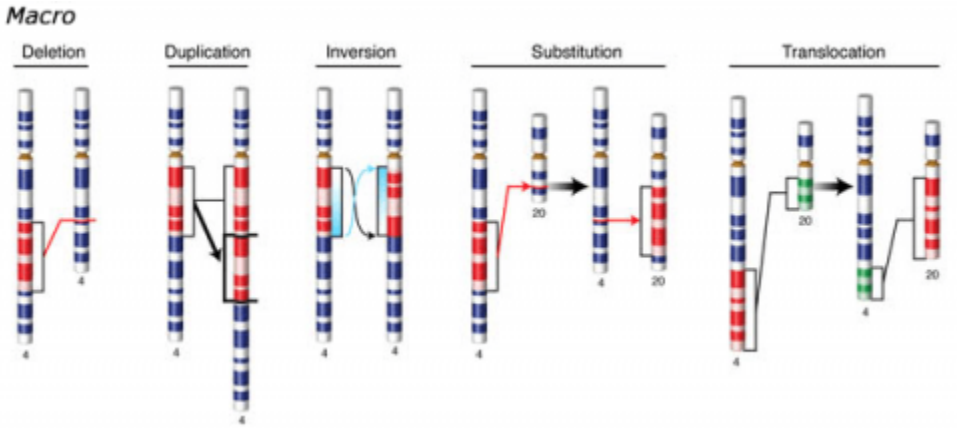


Fig. 5 Basic types of chromosomal mutation. Illustration by NIH-NHGRI, 2011.

- **Chromosomal deletions** are when loss of a chromosome segment occurs.
- **Chromosomal duplications** occur when a chromosome segment is present more than once in a genome or along an individual chromosome. Mutations of this type can involve duplication of chromosome fragments of either noncoding regions or genes that do code for a protein or other gene product. Gene duplications have been important events in the evolution of many crop plants, for example in cotton.Both chromosome deletions and duplications generally result from unequal crossing over during meiosis, whereby one gamete receives a chromosome with a duplicated segment or gene and the other gamete receives a chromosome with a missing or "deleted" segment.
- **Chromosomal inversions** happen when two breaks occur in a chromosome and the broken segment turns 180°—reversing the orientation of the sequence—and then reattaches. Such inverted segments may or may not involve the centromere (termed **pericentric inversion** vs. **paracentric inversion**). A consequence of chromosomal inversions is that they either prevent crossing over or if crossing over occurs, the recombinants may be eliminated during meiosis. During meiosis, inverted chromosome segments may form loops in order to pair with the same (non-inverted) sequence on homologous chromosomes.
- **Chromosomal insertions** (not pictured) and **chromosomal substitutions** are when gain of an extra fragment of chromosome occurs.
- **Chromosomal translocations** entail a change in the location of a chromosome segment. Commonly translocations are **reciprocal** and thus result from exchange of segments between two non-homologous chromosomes.

Transpositions

Chromosome segments can also be translocated to a new location on the same chromosome or to a different chromosome but without reciprocal exchange; both of the latter types of mutations are termed **transpositions**. A **transposon** (also called a **transposable element**) is a DNA element that can move from one location to another. These mobile DNA sequences commonly occur in some genomes and can themselves cause other mutations to occur, depending on where they "transpose".

Discussion

Mutants and mutations are best known in the context of horror films. In the context of plant breeding and more generally crop production, discuss the consequences of mutations—are they good or bad? Which kinds of mutations are desirable and which ones are undesirable?

Sources of Variation

Sources of Genetic Variability

Plant breeding is dependent on differential phenotypic expression. Loci with only one allelic variant (homozygosity) in a breeding population have no effect on the phenotypic variability. Variation can be introduced to breeding populations by various methods:

- **hybridization** and **recombination** by sexual reproduction within or between species or populations
- **genetic transformation** or **genetic engineering** using recombinant DNA methods
- **induced** or **spontaneous mutations** and **transposable elements** (**transposons**)
- chromosome manipulation via change in chromosome number and structure (**ploidization**) [to be discussed in the module on Ploidy—Polyploidy, Aneuploidy, Haploidy]
- **tissue or cell culture** techniques [to be discussed in the module on Ploidy—Polyploidy, Aneuploidy, Haploidy]

CONCEPT OF CROP GENE POOLS

Plant **germplasm** is a term used to refer to an individual, group of individuals or a clone that represents a genotype, population, or species. With reference to a given crop and its wild and cultivated relatives, the concept of **gene pool** (all of the genes shared by individuals in a group of interbreeding individuals) has been applied to categorize a broad range of plant genetic resources according to the ease of gene transfer or gene flow to the particular crop species (Harlan and de Wet 1971).

Gene Pools

Figure 6 depicts three main categories in the original scheme outlined by J. R. Harlan and J.M.J. de Wet (1971) defined as:

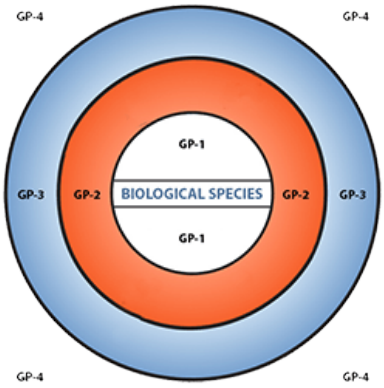
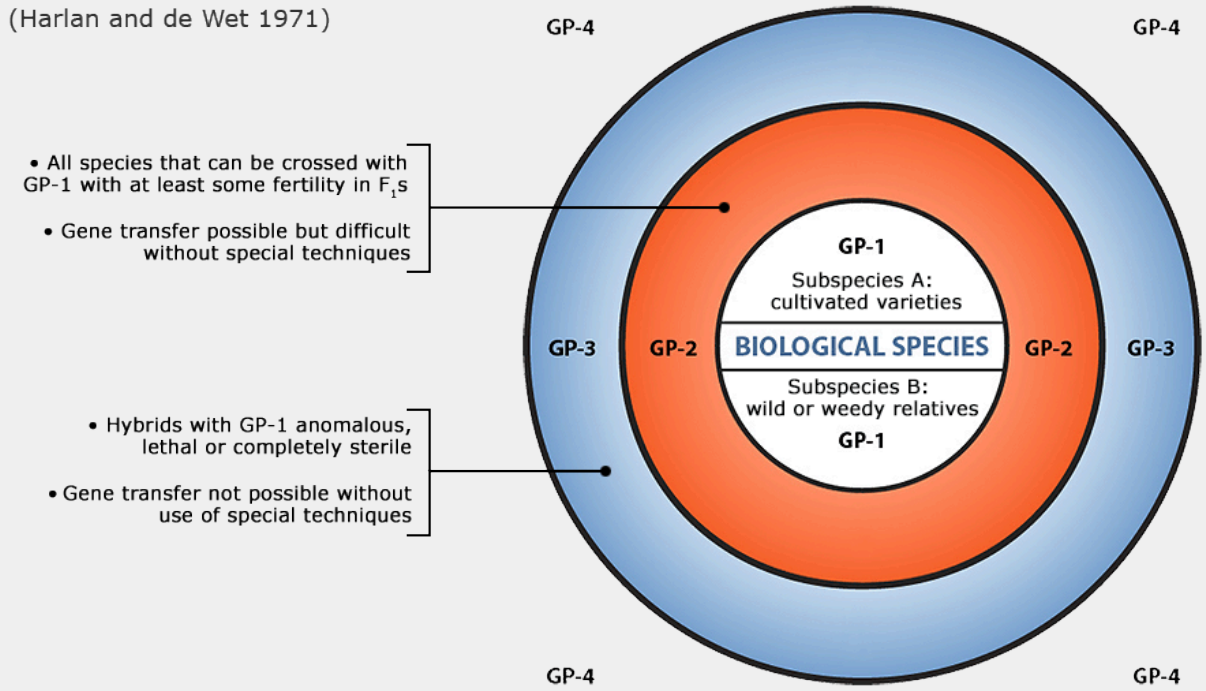


Fig. 6 The modified "gene pool concept". Adapted from Harlan and de Wet, 1971.

- **primary gene pools** (GP-1, consisting of biological species that can be intercrossed easily without problems of fertility in the progeny; including both cultivated varieties and wild progenitors of the crop),
- **secondary gene pools** (GP-2, consisting of more distant relatives that can be intercrossed with difficulty and result in diminished fertility in the hybrids and later generations; including both cultivated and wild relatives of the crop), and
- **tertiary gene pools** (GP-3, consisting of very distant relatives that can be hybridized with the crop only with special techniques, e.g., embryo rescue, due to problems such as sterility, lethality and other abnormalities),

With the addition of a fourth gene pool that contains synthetic variants and lines with nucleic acid sequences that do not normally occur in nature. Methods of genetic engineering relevant to this fourth gene pool category will be covered briefly in the next section of this module.

Gene Pool Concept



Genetic Engineering and Plant Transformation

Genetic engineering, also referred to as **recombinant DNA** or **rDNA** technology or **gene splicing**, involves moving a DNA segment from one organism into another to 'transform' the recipient or host. Through a broad range of techniques encompassing **biotechnology** (for example, gene manipulation, gene transfer, cloning of organisms), novel genetic diversity can be generated that extends beyond species boundaries or can be designed and synthesized *de novo* in molecular laboratories. Potentially, any gene from any species, as well as synthesized segments, could be transferred into a plant using genetic engineering.

Gene transfer can be applied for a variety of objectives:

- Add different or new functions

- Alter existing traits—amplify, suppress, or prevent the expression of a gene already present in the recipient's genome
- 'Tag' and isolate genes in the recipient plant
- Tool for basic gene regulation and developmental studies

Requirements for Gene Transfer

In order to efficiently generate **transformants** (plants that possess DNA introduced via recombinant DNA technologies), the transformation system used must satisfy several requirements.

- Ability to get DNA into host cells in high concentration to increase the probability of incorporation into the host genome
- Incorporation into the host nucleus (or chloroplast if the objective is to minimize gene flow by pollen or to produce a large quantity of therapeutic proteins)
- Integration into host genome (or stabilization as an autonomous replicon—a plasmid or minichromosome)
- The introduced gene is expressed and translated properly

GENERAL STEPS

1. Identification and isolation of a gene that confers a desired trait.
2. Introduce the gene into a suitable construct and carrier, such as plasmids or bacterial vectors, for delivery into the host.
3. Introduce the DNA into the host.
4. Identify and select transformants.
5. Regenerate plants.
6. Assay for expression of the trait.
7. Test for normal sexual transmission, or asexual propagation, of the transferred gene.

Construct and Carrier

The carrier that will deliver the DNA into the host should have certain features.

- Sites in which to insert passenger DNA sequences (gene of interest, plus a selectable marker gene if the gene of interest does not allow for easy selection of transgenic plants).
- Sequences to mediate integration into the host genome
- Selectable marker gene for identification and selection of transformants

Usually, the DNA sequence to be transferred into the host is joined with other sequences to facilitate transfer, incorporation, and expression of the gene. Here's a generalized construct for a T-DNA vector, a carrier derived from an Agrobacterium tumefaciens plasmid.

Introduce the DNA

Several methods are available to introduce the DNA into the host.

- Vector-mediated transfer
- Direct DNA uptake—DNA cannot be taken directly into cells having a cell wall so protoplast must be used.
- Microinjection—DNA is injected directly into the host nucleus
- Acceleration of DNA—coated particles – particles are "shot" into the cell (particle bombardment or gene gun)

Genetic transformation plays an important role in modern-day crop improvement. The first transgenic plant was created in 1983. By 1996, there were already 1.7 million hectares of genetically modified (GM) crops and this number increased 100-fold to 170 million by 2012 and is still increasing. The majority of the GM crops (soybean, corn, cotton, papaya, canola and sugarbeet) were created by the use of *Agrobacterium tumefaciens* (vector-mediated transfer) to resist either herbicides or insects. Herbicide-resistant crops greatly simplified weed management where mechanization in agriculture is high. Insect-resistant crop plants produce stable yields. The tremendous expansion of GM crop production, however, is not realized without controversies. There is currently an intense public debate over the impact of GM crops

on human and animal health. Besides health issues, other concerns surrounding GM crops are whether they can create superweeds by crossing to related weeds, become invasive or cause unintended harm to wildlife.

Bt Gene: Vector Constructs

Let's follow the transfer of a Bt gene into a plant.

The Bt gene was identified and isolated from *Bacillus thuringiensis*, a bacterium. The gene produces a protein that has toxic effects on Diptera (flies), Lepidoptera (butterflies and moths), and Coleoptera (beetles) species. Soybean was transformed using *Agrobacterium tumefaciens*. Two different T-DNA vector constructs carrying the Bt gene and one control construct were tested for effectiveness in transforming soybean cells and expressing the Bt toxin (Fig. 7).

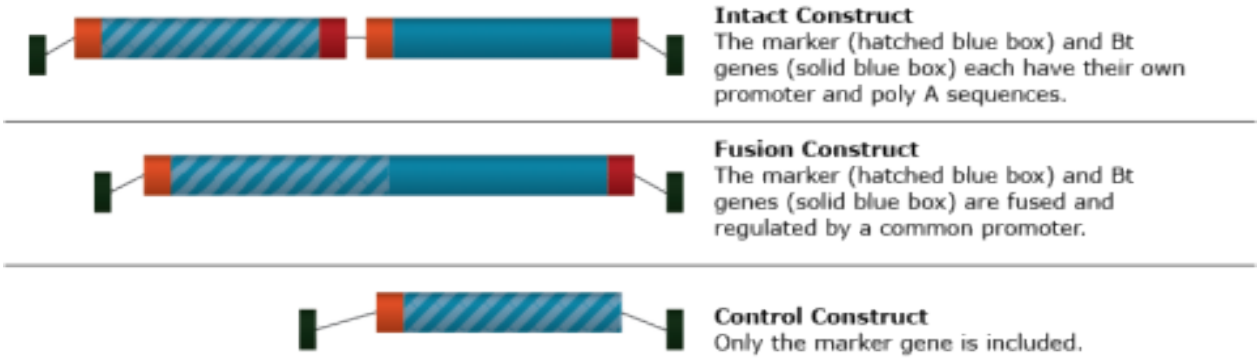


Fig. 7 Two types of T-DNA vector constructs and control construct.

Bt Gene: Transformation

Leaf disks from soybean plants were infected and cultured on selective (+) and control (-) media—the selective medium was gradually enriched with an antibiotic (Fig. 8). At time 1, the leaf disks were infected with the respective constructs. Conditions were used to promote callus production and growth. At time 2, the plates were evaluated for calli formation.

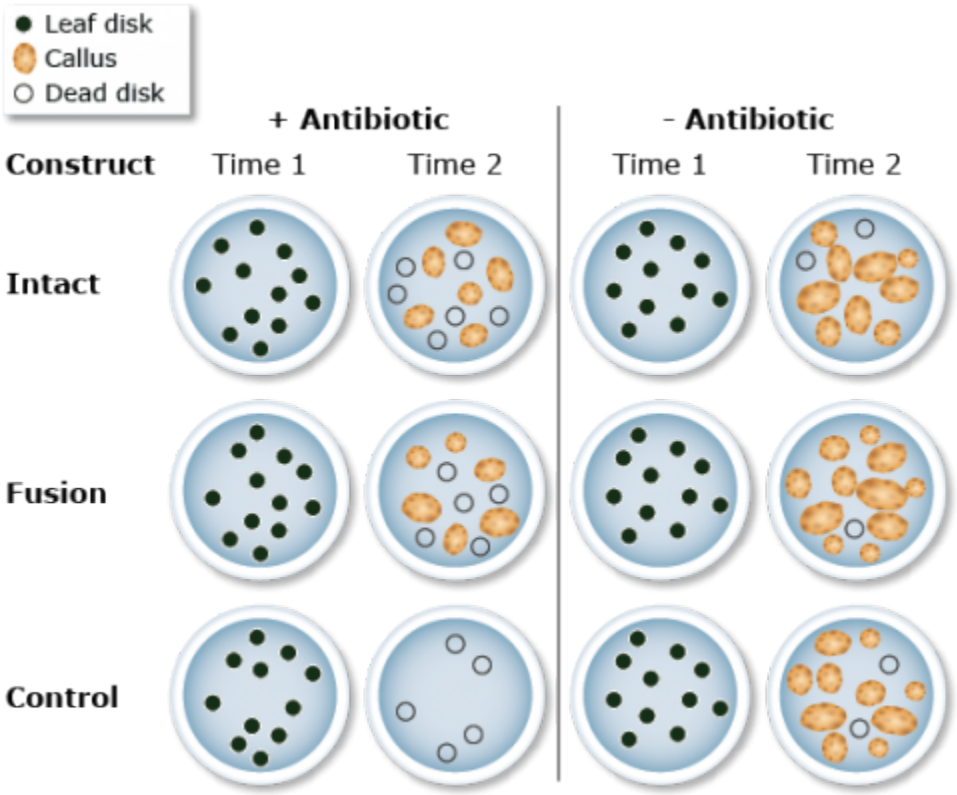


Fig. 8 Leaf disks infected and cultured to promote callus formation.

Transformed cells are able to develop into calli. These are selected and transferred to a medium containing both antibiotics and growth regulators that promote the formation of shoots and roots (Fig. 9).

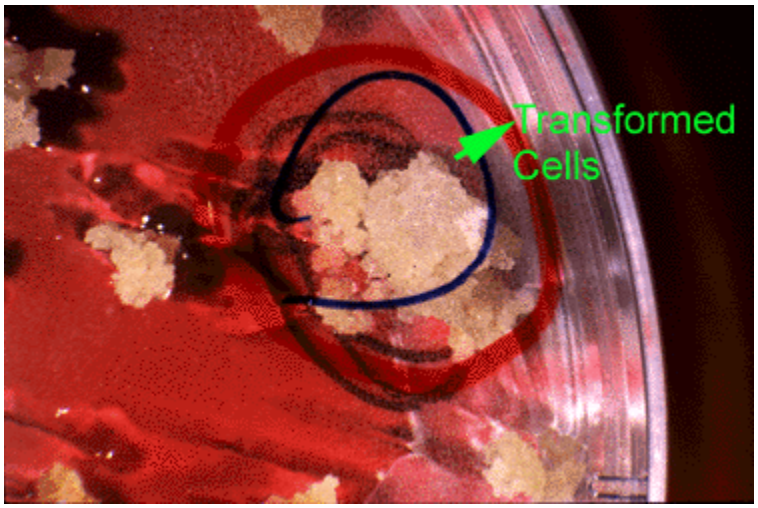


Fig. 9 Transformed cells proliferate. Photo by Bronwyn Frame, ISU Plant Transformation Facility.

Bt Gene: Insect Resistance Evaluation

Plantlets are regenerated and transferred to pots containing sterilized soil. Nearly all of the regenerated plants exhibited normal morphology and vigor. A few had chlorophyll deficiencies— these were eliminated from the study. The remaining regenerated plants were evaluated for insect resistance. An equal number of insect larvae are placed on each regenerated plant; plants are isolated to prevent insects from moving among plants.

Fig. 10 Insect damaged soybean plant. Photo by Iowa State University.

Fig. 11 Healthy soybean plants. Photo licensed under public domain via Wikimedia Commons.

Bt Gene: Insect Resistance Data Analysis

After several days, the dead larvae on each plant are counted. Here are the data.

Table 1 Insecticidal activity as indicated by the number of plants exhibiting different levels of larvae mortality. The total number of plants regenerated from each construct treatment was unequal.

Gene Construct	No. of Plants Showing % Larvae Mortality			
	< 25 %	25% – 50%	51% – 75%	76% – 100%
Intact	14	7	0	0
Fusion	2	6	2	20
Control	29	0	0	0

Advantages and Disadvantages

Although recombinant DNA technologies have some problems, the technologies offer several advantages.

Advantages	Disadvantages
Characters can be transferred from divergent species without the limitation of sexual compatibility	Difficult to identify and isolate gene
Single gene or gene sets can be transferred into important breeding lines without the deleterious effects of linked genes.	**Insertion is random**
Transferred character ordinarily exhibits **dominant, single-gene inheritance**	Difficult to obtain **proper expression**
	Still must screen at whole plant level and under normal production conditions
	Expensive

Despite its limitations, plant transformation has additional advantages over conventional breeding. Millions of cells, with regeneration capacity, can be screened for the desired trait in a few weeks. A desired gene can be transferred without the necessity of generations of breeding to move the trait from one line into another. Recombinant DNA technology can also be used to place synthetic genes into plant genomes.

Insertion is Random

The insertion point of the transferred gene cannot be controlled. Plants contain an estimated 500,000 to 5,000,000 kilobases (kb) of DNA. The maize genome, for example, has more than 4,000,000 kb. Generally, transferred DNA involves relatively small amounts of DNA, on the order of 10 kb, so the insertion ordinarily has little effect on chromosome pairing, recombination, or mitosis. The incorporated gene may or may not affect other genes in the recipient genome, depending on where it inserts.

- In a non-coding region—no effect on the recipient genome

- In a gene that occurs in multiple copies in the genome—the effect, if any, is usually not detectable.
- In a single-copy gene—inactivates or alters the expression of the single-copy gene.

Insertion into a single copy gene is rare. If it does insert into a single-copy gene, inactivation or alteration of the expression of the disrupted gene may be undetected, or may cause favorable or adverse results—albino and other chlorophyll deficiencies are common problems.

Dominant, Single-gene Inheritance

Why do these characters generally exhibit dominant, single-gene inheritance? Traits acquired via gene transfer often add a function to the transformed plant. Because the transferred trait is unique in the transformant's genome, the transformant does not possess any contrasting alleles for the character. Thus, its inheritance is expected to be dominant and as a single gene.

Gene transfer is also used to suppress or eliminate, or to amplify the expression of genes already possessed by the host plant. These are also usually designed to behave as dominants.

Proper Expression

Expression involves many steps, all of which must occur properly to obtain the desired phenotype. Expression must be appropriately regulated.

- Generation of the gene's product-requires proper transcription, mRNA processing, and translation.
- Location of expression must be in the appropriate plant part.
- Timing of expression needs to occur at the right stage of the plant's development.
- Amount needs to be at an effective level or extent of expression to generate the desired

phenotype.

For Your Information

Table of Codons

Amino acids are the building blocks of **polypeptides**, **proteins**, and **enzymes**. The order of the **nucleotides** on a strand of RNA, as transcribed from **DNA**, determines the order of amino acids in a polypeptide. Each group of three consecutive nucleotides of the RNA codes for a particular amino acid, or the beginning or end of the message. These triplets of nucleotides are called **codons** (Fig. 20).

Fig. 20 A codon is a trinucleotide sequence of DNA or RNA that corresponds to a specific amino acid. Illustration from NIH-NHGRI, 2011.

The genetic code, or instructions from a gene that direct the cell to make a specific protein, is usually based on the messenger RNA (mRNA) sequence (Fig. 21). In mRNA, uracil (U), rather than thymine (T), is the nucleotide base that complements adenine (A) on the DNA strand; guanine (G) complements cytosine (C) in both DNA and RNA.

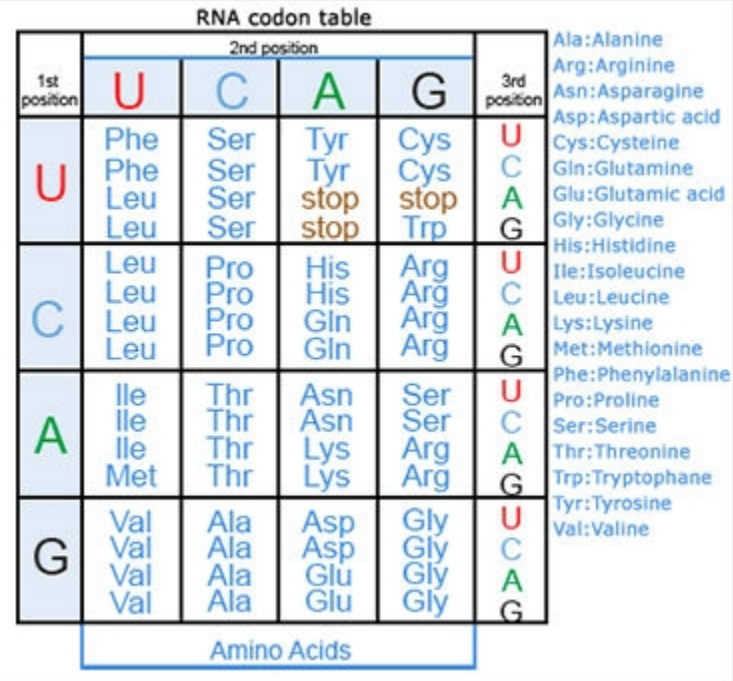


Fig. 21 Genetic code is based on the nucleotide sequence on the mRNA. Illustration from NIH-NHGRI, 2011.

Here are some examples of codons:

- A-A-A and A-A-G signal the amino acid lysine (Lys)
- G-A-A and G-A-G code for glutamine (Gln)
- A-U-G signals the start of a coding sequencing and codes for methionine (Met)
- U-A-A, U-A-G, and U-G-A are stop codons.

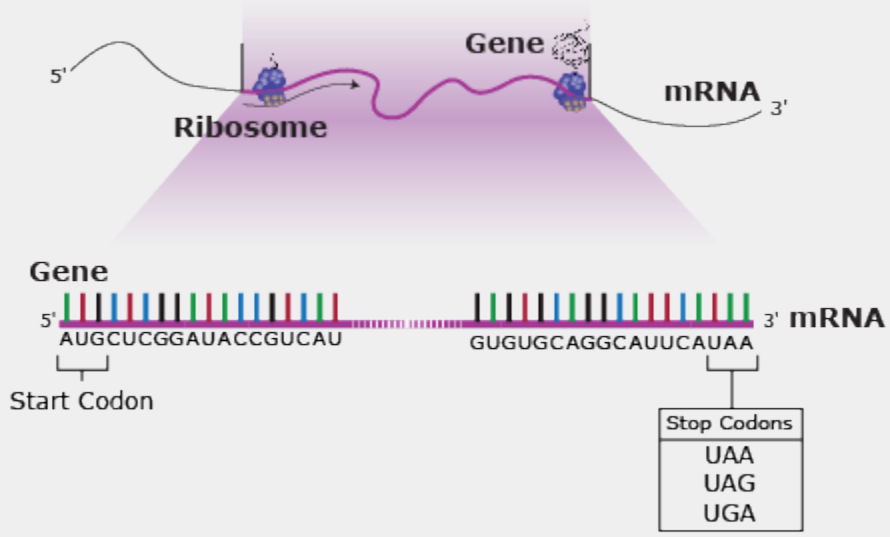


Fig. 22 Stop codons signal a halt to protein synthesis. Illustration from NIH-NHGRI, 2011.

Of the 64 possible combinations of three bases, 61 specify an amino acid, while the remaining three combinations are **stop codons**, or trinucleotide sequences that indicate the end of the message, terminate translation of that mRNA section, and signal a "stop" to protein synthesis (Fig. 22).

The portion of a DNA molecule that, when translated into amino acids, contains no stop codons is referred to as an **open reading frame** (Fig. 23).

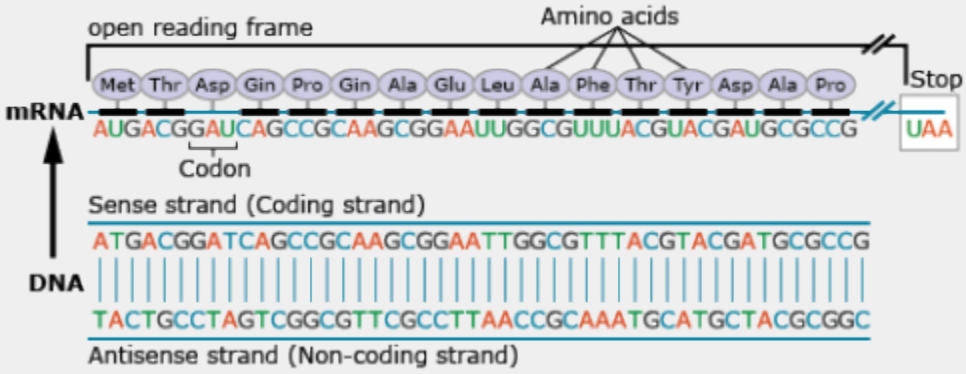


Fig. 23 An open reading frame is a fragment of DNA that when translated contains no stop codons. Illustration from NIH-NHGRI, 2011.

References

Collard, B.C.Y., M.Z.Z. Jahufer, J.B. Brouwer, and E.C.K. Pang. 2005. An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. Euphytica 142: 169-196.

Falconer, D.S. and T.F.C. Mackay. 1996. Introduction to Quantitative Genetics. 4th edition. Longman Pub. Group, Essex, England.

Harlan, J.R., and J.M.J de Wet. 1971. Towards a rational classification of cultivated plants. Taxon 20: 509-517.

Gepts, P. 2002. A comparison between crop domestication, classical plant breeding, and genetic engineering. Crop Science 42: 1780-1790.

Nageswara-Rao, M. and J.R. Soneji. 2008. Molecular genetic markers: what? why? which one for exploring genetic diversity? Perspectives-The Science Advisory Board, 8 September 2008 [available online September 23, 2011, http://www.scienceboard.net/community/perspectives.210.html

Neuffer, M.G, E.H. Coe, and S.R. Wessler. 1997. Mutants of Maize. Cold Springs Harbor Laboratory Press.

NIH-NHGRI (National Institutes of Health. National Human Genome Research Institute). 2011. Talking Glossary of Genetic Terms. [available online February 6, 2017, http://www.genome.gov/glossary/

How to cite this chapter: Merrick, L., A. Campbell, D. Muenchrath, and S. Fei. 2016. Mutations and Variation. In W. P. Suza, & K. R. Lamkey (Eds.), *Crop Genetics*. Iowa State University Digital Press. DOI: 10.31274/isudp.2023.130

Ploidy: Polyploidy, Aneuploidy, and Haploidy

Deborah Muenchrath; Arden Campbell; Laura Merrick; Thomas Lübberstedt; and Shui-Zhang Fei

Introduction

Not all plant species are **diploids**. In fact, 75% of all angiosperms are **polyploids**, which are characterized by having more than two sets of chromosomes in their somatic cells. About 70% of undomesticated grasses and 25% of legumes are polyploids. Among domesticated crop species in general, 75% have been found to be polyploid, with roughly similar percentages found for both annual and perennial species. Major crops, such as wheat, alfalfa, potato, cotton, and sugarcane, are polyploids. There are also plants that do not possess complete sets of chromosomes. Aneuploids have abnormal numbers of chromosomes and vary by the addition or deletion of specific individual chromosomes that otherwise would be present in the normal crop genome. Ploidy reduction produces **haploids**, which have only a single set of homologous chromosomes instead of the pair found in their diploid counterparts. Haploid plants are very valuable in certain breeding applications.

The number of chromosome sets possessed by a crop influences its genetics and thus, the strategies applied for its improvement. Plant breeders can alter chromosome numbers to modify and exploit genetic variability.

Learning Objectives

- Identify types and characteristics of polyploidy, aneuploidy, and haploidy.
- Understand how plants that vary in ploidy level occur naturally or are created artificially.
- Become familiar with the genetic behavior of polyploids, **Aneuploids**, haploids, and implications of changes in chromosome number for reproduction and productivity.
- Learn about the strengths and weaknesses of variation in ploidy level in plant breeding and genetic applications.
- Explore tissue and cell culture methods and their applications.

Concepts for Polyploidy

Ploidy

Polyploidy can be rather complex. The terminology and symbols used in connection with polyploidy communicate much information—understanding these will help clarify and simplify an otherwise complex topic.

- **Ploidy** refers to the number of chromosome sets in a cell. Prefixes are used to specify the number of chromosome sets in a particular organism. The symbol x is used to indicate the number of chromosomes in a set. Monoploids have one set (1x) and diploids have two sets (2x) of chromosomes, and so forth.

Table 1 Corresponding number of chromosomes sets to ploidy level.

Ploidy Level	Number of Chromosome Sets
Monoploid	1x
Diploid	2x
Triploid	3x
Tetraploid	4x
Pentaploid	5x
Hexaploid	6x
Heptaploid	7x
Octoploid	8x
Nonaploid	9x
Decaploid	10x

Polyploid

- **Polyploid** is a general term indicating multiple (more than two) sets of chromosomes.
- **Genome** is a set of chromosomes that is inherited together, assuming normal meiosis and mitosis. Each genome is composed of the basic chromosome number, x, and each kind of

chromosome is represented only once in each set.

Several symbols are used as shorthand to convey additional information about the chromosome constitution of the species or individual.

- **n** — indicates the haploid or gametic chromosome number of the species. (Note: **n** does not tell anything about the number of chromosome sets).
- **2n** — denotes the diploid or somatic number of the species.
- **x** — indicates the basic number of chromosomes in each set or genome.

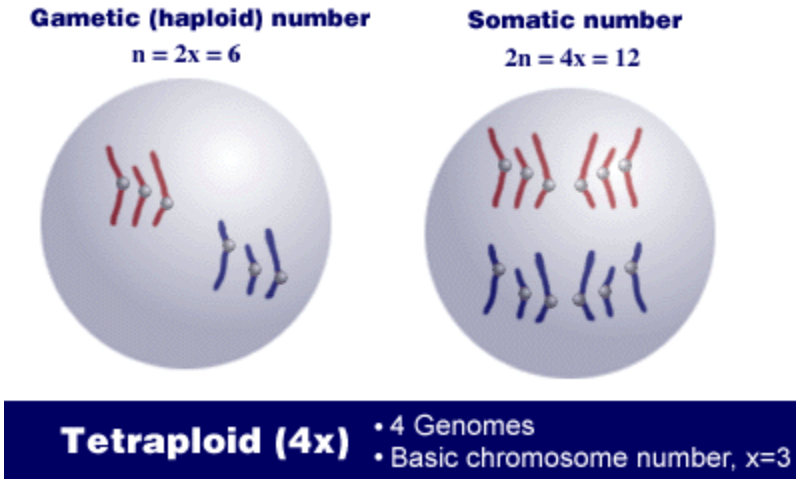


Fig. 1 (Source: NIH-NHGRI)

For example, a **tetraploid**, 4x, has 4 genomes or sets of chromosomes. The number of chromosomes in its somatic cells is 2n = 4x, and the number in its gametes is n = 2x. In this illustration, let three be the basic chromosome number, so x = 3. This means that each genome or chromosome set consists of 3 chromosomes.

Types of Polyploidy

Euploidy

There are two general types of ploidy, which include plants that have either one or more complete sets of chromosomes present in their genome (euploids) or those that have partial

sets due to the absence of at least one of their individual chromosomes or presence of at least one extra one (aneuploids):

Euploidy refers to the number of chromosome sets in a cell. Prefixes are used to specify the number of chromosome sets in a particular organism. The symbol x is used to indicate the number of chromosomes in a set. Monoploids have one set (1x) and diploids have two sets (2x) of chromosomes, and so forth.

- **Haploidy**—individual with half (n=x) of the somatic cell chromosome number.
- **Diploidy**—individual with two sets of the basic, complete genome (2n=2x).
- **Polyploidy**—individual with more than two basic, complete sets of chromosomes in its somatic cells. Among polyploids, there are two main types:
  - Autoploidy—individual has more than two complete chromosome sets from a single genome. Also known as autopolyploidy.
  - Alloploidy—individual has two or more genomes contributed from different parental species in their ancestral lineage. Also known as allopolyploidy. Most naturally occurring polyploids are alloploids. **Amphidiploids** (also called amphiploids) are **allotetraploids** that contain two sets of genomes from each of two different parents. Amphidiploids are so-called because they behave like diploids during meiosis.

Autoploidy

Autoploids commonly occur as a result of duplication of the genome(s) of a single species, thus, the genomes possessed by an autoploid are identical. An autotetraploid (4x) has four sets of the same genome, e.g., four sets of the A genome (AAAA). The following are characteristics associated with autoploids:

- Greater ability to colonize new habitats than diploid ancestors—Due to gene buffering, autoploids often show a slower response to selection, but more adaptive potential.
- Dosage effect of gene expression—Additive effect of the alleles increases the number of phenotypes. There is a linear relationship between gene expression and number of gene copies.
- Larger cells and nuclei compared to their respective diploid counterparts—Because of the greater cell size, autoploids tend to have greater vegetative growth and produce larger structures. This feature has been utilized by breeders to increase yield of crops harvested for their vegetative structures.

- Reduced fertility—Abnormalities in meiosis may interfere with chromosome pairing, resulting in unbalanced chromosomal distribution between daughter cells, and thus, nonviable gametes. Gametes possessing an extra chromosome or missing a chromosome (i.e., aneuploids) are usually nonviable. This feature is utilized by breeders to develop seedless crops such as seedless watermelon.

Alloploidy

Alloploids arise when the genomes of two or more unrelated species are combined in a single individual.

Several characteristics are associated with alloploidy:

- Broadened genetic base. Alloploids (also known as amphiploids) behave like diploids and result in new species.
- Increased allele diversity and heterozygousity—Increase in possible allele combinations can provide expanded opportunities for breeding.
- Novel phenotypic variation—Genome interactions and changes in gene expression can occur in newly synthesized alloploids. Such gene changes can include transfer of sequences between genomes and gene conversion, loss, or silencing.
- **Sterile** unless genomes are doubled and chromosomes pair correctly. Chromosome doubling also occurs spontaneously in nature, mainly through the fusion of unreduced gametes.

Alloploidy has impacted the evolution of some major crops, such as wheat, cotton, tobacco, and various *Brassicas* (mustards). Information about the ploidy relationships between certain species has improved our understanding of the genetic origin of polyploid crops and facilitated their improvement.

In addition to identifying the genetic origin of a crop, genes, as well as whole genomes, can be transferred among related species to obtain novel genotypes and to combine the favorable qualities of different species as is demonstrated in the creation of Triticale.

Colchicine

Colchicine is a chemical that is commonly used to artificially double chromosomes. Colchicine is a toxic natural product that is found in the bulbs of autumn crocus (*Colchicum autumnale*). It is used as a medicine to treat the human disease gout, but it is also used to induce polyploidy in plants. Originally it was extracted from crocus bulbs, but is now also manufactured synthetically. When plant seeds or meristem tissues (apical tip, shoots, or suckers) are soaked in colchicine, it makes the cell walls permeable and prevents normal chromosome division.

Colchicine functions as a disruptor of mitosis by inhibiting microtubule formation, thus preventing normal chromosome segregation to occur. After seed treatment with the chemical, half the gametes resulting from meiosis contain no chromosomes, while the other half contains diploid instead of the normal haploid number. This disruption leads to embryos with double the usual number of chromosomes. Colchicine is used to make infertile hybrids or haploid plants fertile by restoring doubled chromosomes. The latter plants are known as **doubled haploids**. An example of the former is development of hexaploid and octaploid triticale.

Colchicine is also used in the study of **karyotypes**, which refer to the chromosome constitution of the cell (specifically, the number and morphological appearance of chromosomes in the nucleus). In order to observe chromosomes under a light microscope, cells are treated with the chemical near the middle of mitosis during metaphase—a point in the cell cycle when chromosomes are most dense and therefore most visible. Colchicine treatment arrests these chromosomes by preventing the formation of the spindle microtubules.

Wheat

Wheat is an example of a crop whose origin has been extensively studied and is a classic example of a polyploid crop that evolved through the combination of genomes from related species. Wheat has three sets of chromosomes, designated as A, B, and D. Each of these sets contributes 7 chromosomes.

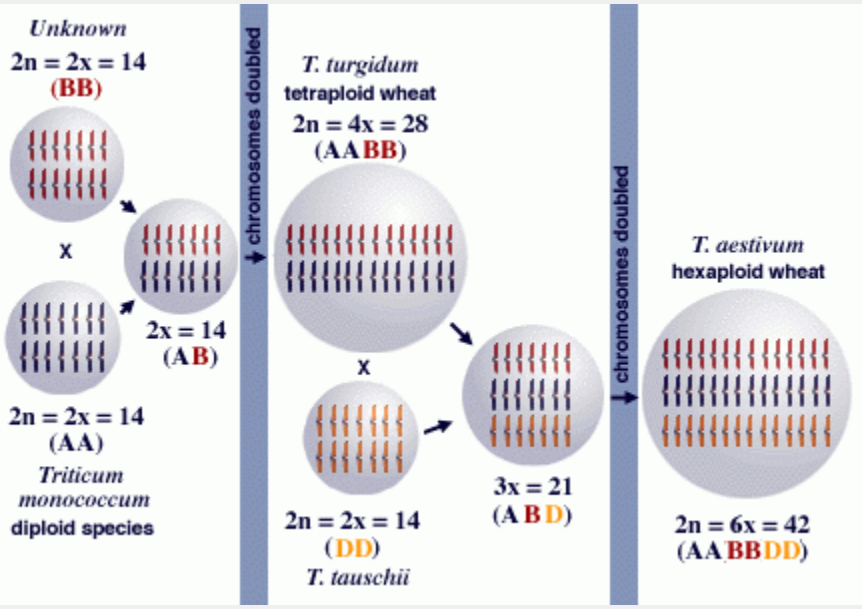


Fig. 2

- wheat = AABBDD
- 1A, 2A, ..., 7A;
- 1B, 2B, ..., 7B;
- 1D, 2D, ..., 7D.

Brassica Evolution—Triangle of U

The "Triangle of U" is a theory that the genomes of three ancestral species of *Brassica* combined through interspecific hybridization. *Brassica* species include both vegetable and oilseed crops, such as cabbage, broccoli, rapeseed, and various mustards.

The original theory was developed by Woo Jang-choon, a Korean botanist who published under the Japanese name Nagaharu U. He made artificial hybrids between closely related, but distinct diploid and polyploid species and examined how chromosomes paired in the progeny. Woo's theory was later confirmed by examination of DNA from the group of related *Brassica* species.

The different colors in the chromosomes depicted here represent the three ancestral genomes:

Genomes of three diploid species

AA: 2n = 2x = 20—*Brassica rapa* (syn. *Brassica campestris*) – turnip, Chinese cabbage

BB: 2n = 2x = 16—*Brassica nigra* – Black mustard

CC: 2n = 2x = 18—*Brassica oleracea* – cabbage, kale, broccoli, Brussel sprouts, cauliflower

AABB: 2n = 4x = 36—*Brassica juncea*—Indian mustard

AACC: 2n = 4x = 38—*Brassica napus*—Rapeseed, rutabaga

BBCC: 2n = 4x = 34—*Brassica carinata*—Ethiopian mustard

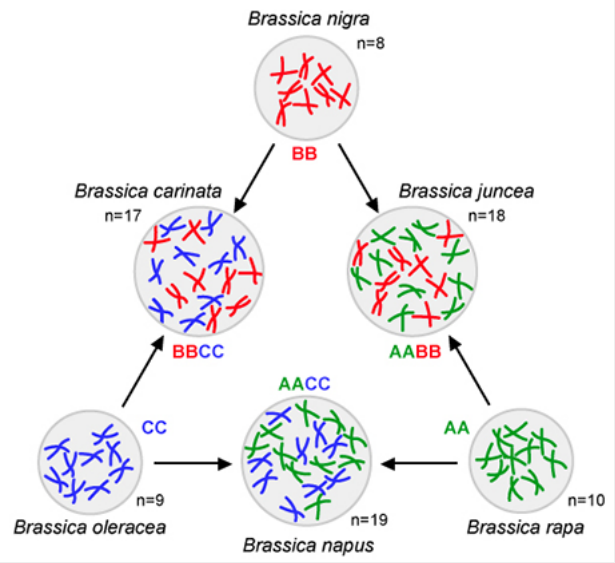


Fig. 3 The "Triangle of U": evolution and relationships among Brassica species. Adapted from U, 1935.

Autoploids and Alloploids

Many important crops are naturally occurring polyploids—either autoploids or alloploids. Letters are often used to indicate to the different source species that have been identified as

contributors to crop genomes. However the same letter does not denote the same genome unless the species are closely related evolutionarily. For example, tall fescue (an alloploid) and perennial ryegrass (a diploid) both share the L genome in common. Pima and Upland cotton (both alloploids) share A and D genomes in common, but only share the A genome with two other cultivated cottons, Levant and Tree cotton (both diploids).

Other polyploid crops include sugarcane, yam, white potato, orchardgrass, highbush blueberry, persimmon, basil, nutmeg, and some types of asparagus and celery. Many horticultural species of flowers and ornamentals are polyploids. Some of these polyploids occur naturally and others were manipulated into polyploids through plant breeding efforts. Two examples of artificially created alloploids that are produced commercially include the cereal grain triticale (*Tritcosecale*) and the vegetable crop known as radicole (*Raphanobrassica*). Both were developed by intergeneric hybridization among closely related crop genera: wheat-rye and radish-cabbage (or radish plus other *Brassicas*).

Recent studies involving molecular techniques have revealed that many diploid crops with high chromosome numbers (2n>18-20) are actually **paleopolyploids**—plants whose genomes result from ancient genome duplications that occurred at least several million years ago. Known paleopolyploids of either alloploid or autoploid origins include such crop species as soybean, squash, maize, and lettuce.

Diploidization

In a process known as **diploidization** over time polyploid genomes may return to diploid status, although certain genes may be retained in multiple copies. Conversely, the term **polyploidization** refers to cycles of hybridization and chromosome doubling that result in either allopolyploids or autopolyploids.

Table 2 Examples of naturally-occurring euploid crops.

Crop Type	Diploid or Polyploid	Genome	Ploidy (x)
African rice	Diploid	A'A'	2x
Asian rice	Diploid	AA	2x
Bread wheat	Alloploid	AABBDD	6x
Durum wheat	Alloploid	AABB	4x
Rye	Diploid	RR	2x
Oats	Alloploid	AACCDD	6x
Broccoli, Cabbage, Kale, Cauliflower, Brussel sprouts	Diploid	CC	2x
Turnip, Chinese cabbage	Diploid	AA	2x
Rapeseed, Rutabaga	Alloploid	AACC	4x
Black Mustard	Diploid	BB	2x
Brown Mustard	Alloploid	AABB	4x
Ethopian Mustard	Alloploid	BBCC	4x
Radish	Diploid	RR	2x
Radicole	Alloploid	RRCC, RRAA	4x
Sweet potato	Autoploid	BBBB, BBBBBB	4x, 6x
Pima cotton	Alloploid	AADD	4x
Tree cotton	Diploid	AADD	4x
Upland cotton	Alloploid	AADD	4x
Levant cotton	Diploid	AA	2x
Safflower	Diploid	BB	2x
Alfalfa	Autoploid	SSSS	4x
Bermuda grass	Diploid, Autoploid	AA, AAAA	2x, 4x
Tall fescue	Alloploid	PPLLXX	6x
Perenial rye grass	Diploid	LL	2x
Timothy	Alloploid	AAAABB	6x

Crop Type	Diploid or Polyploid	Genome	Ploidy (x)
Banana	Diploid, Alloploid, Autoploid	AAA, AAA, ABB, AB, AAAB, AABB, ABBB	2x-4x
Plum, European	Alloploid	CCSSSS	6x
Coffee	Alloploid	AABB	4x
Tobacco	Alloploid	SSTT	4x

Table 3 Examples of artificially developed alloploid crops.

Crop Type	Hybrid or Parent	Genome
Triticale	Intergeneric hybrid	AABBRR, AABBDDRR
Bread wheat	Parental species	AABBDD
Durum wheat	Parental species	AABB
Rye	Parental species	RR
Radicole	Intergeneric hybrid	RRCC, RRAA
Broccoli, Cabbage, Kale, Cauliflower, Brussel sprouts	Parental species	CC
Turnip, Chinese cabbage	Parental species	AA
Turnip, Chinese cabbage	Parental species	RR

Aneuploidy

Aneuploidy — organism has a partial set of chromosomes due to addition or deletion of specific chromosome(s). or one or more chromosome sets. It is a state in which the number of chromosomes is not an exact multiple of the haploid number.

There are numerous possible kinds of aneuploidy.

Types of Aneuploids

Aneuploids are described relative to the euploid condition. The general genomic formula for aneuploids is somatic chromosome number plus or minus the number of extra or missing chromosomes relative to the euploid somatic number: 2n ± one or more chromosomes. The following table describes the classification system used for aneuploids which includes not only the addition or absence of individual chromosomes (as described in the table below), but also includes the presence or absence of chromosomes characterized by partial rearrangements.

Kind of Aneuploid	Genomic Formula	Interpretation Relative to Euploid Condition
Nullisomic	2n – 2	Euploid minus one pair of chromosomes
Monosomic	2n – 1	Euploid minus one chromosome
Trisomic	2n + 1	Euploid plus one chromosome
Tetrasomic	2n + 2	Euploid plus one pair of chromosomes

How has aneuploidy been useful in plant breeding? Breeders and geneticists use aneuploids as a tool to

- identify the chromosomal location of specific genes, or
- substitute a particular chromosome into a genotype.

Chromosome transfer or substitution provides an efficient method to transfer specific characters without accompanying adverse traits. It also facilitates chromosome transfer from alien species, enabling the recipient population to acquire genes that are unavailable within its own gene pool. Aneuploids can be used to increase genetic diversity.

Induction of Polyploids, Aneuploids, and Haploids

Natural Induction of Polyploids

Polyploidy occurs spontaneously in nature. There are two mechanisms for natural induction of polyploidy.

1. **Unreduced gametes**—When the chromosome number fails to be reduced during meiosis, unreduced gametes result and can produce polyploids if the unreduced gamete is successfully fused with reduced or unreduced gametes of the opposite sex.

Table 4

Female Gamete X	Male Gamete ⇒	Resulting Zygote
n (reduced)	n (reduced)	2n (normal)
2n (unreduced)	n (reduced)	2n + n (1 additional set of chromosomes from female)
n (reduced)	2n (unreduced)	n + 2n (1 additional set of chromosomes from male)
2n (unreduced)	2n (unreduced)	2n + 2n (2 additional sets of chromosomes, with 1 additional set from each parent)

2. **Somatic chromosome doubling**—Normal gamete fusion occurs but the chromosomes spontaneously double in the zygote or other somatic tissue. Such chromosome doubling is rare.

Ordinarily, conventional crosses between parents having different genomes (referred to as a "wide cross") are unsuccessful. Their offspring usually exhibit:

- Reduced fertility as a result of problems in chromosome pairing during cell division.
- Reduced seed set.
- Reduced viability associated with genomic instability, especially in those with an odd number of chromosome sets (e.g., triploids, pentaploids).

Chromosome doubling generally restores or improves genetic stability, viability and fertility.

Artificial Induction of Polyploids

Polyploidy can also be induced artificially. Chemicals such as colchicine or environmental shocks can be used to disrupt normal chromosome division and induce polyploidy. **Meristematic** tissues are especially susceptible to such disruptions.

Induction of Aneuploids

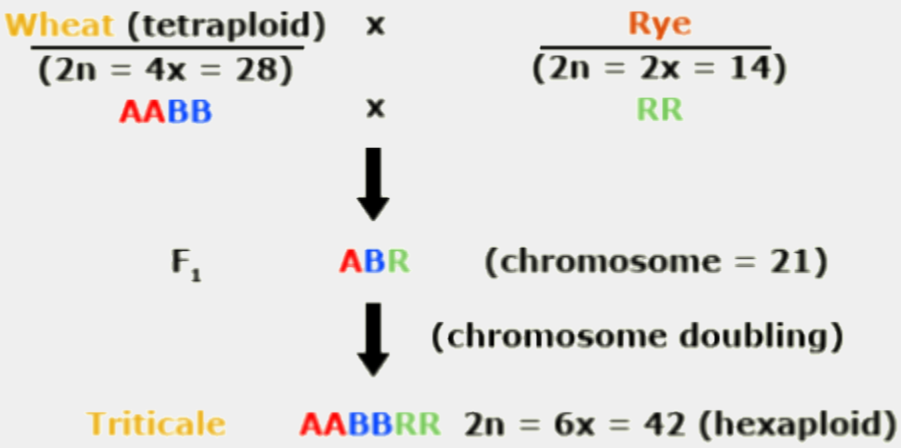
Aneuploidy occurs during cell division when chromosomes do not segregate properly into daughter cells. When meiosis occurs, germ cells divide to create male and female gametes, typically with each sperm and egg ending up with the same number of chromosomes. However, when nondisjunction occurs, an extra copy may end up in one gamete and the other may not have any copies. In animals, aneuploidy is usually fatal. Plants can tolerate higher levels of aneuploidy, but aneuploidy female gametes are typically more viable than male gametes.

Triticale

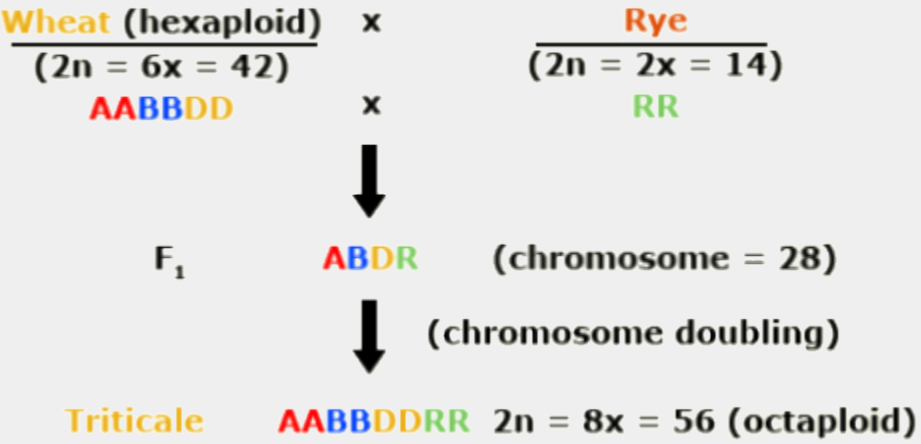
Triticale is a combination of wheat and rye. Efforts to mate these two crops began in 1888. Why was triticale developed? Triticale (*Triticosecale*) merges the desirable qualities of wheat and rye.

- Wheat possesses favorable flour and baking qualities and flavor.
- Rye contributes to improved vigor. Rye is winter hardy, adaptable to poorer soils and growing conditions, and resistant to some diseases.

Hexaploid triticale was created by combining the genomes of a tetraploid wheat (Triticum turgidum) with rye (Secale cereale). Rye is diploid. In this schematic, the wheat genomes are represented as A and B and the rye genome by R.



Octaploid triticale was formed through combination of hexaploid wheat (*Triticum aestivum*) and rye. The additional wheat genomes are represented by D in this illustration.



In the development of hexaploid and octaploid triticale, colchicine was used to double the chromosome number in the newly synthesized interspecific hybrids. Duplication of the genome resulted in restored fertility due to normal pairing of chromosomes in meiosis. With subsequent breeding, triticale was further improved. Today it resembles wheat, except that it has more vigorous growth and produces larger spikes and kernels.

Induction of Haploids

Spontaneous haploid plants occur at low frequency in many species. In maize, for example, it occurs in the order of less than 1:1000. Haploid plants are of great interest in plant breeding, because genome doubling of haploids by chemical treatment (e.g., by use of **colchicine**) is possible, which leads to **doubled haploid** (**DH**) plants and lines, in which all loci are homozygous. Thus the development of inbred lines can be achieved in a much shorter time than by continued self-pollination starting from a heterozygous plant. To efficiently use DH lines, the spontaneous rates of haploid induction are not sufficient. Thus, procedures have been developed to induce haploidy.

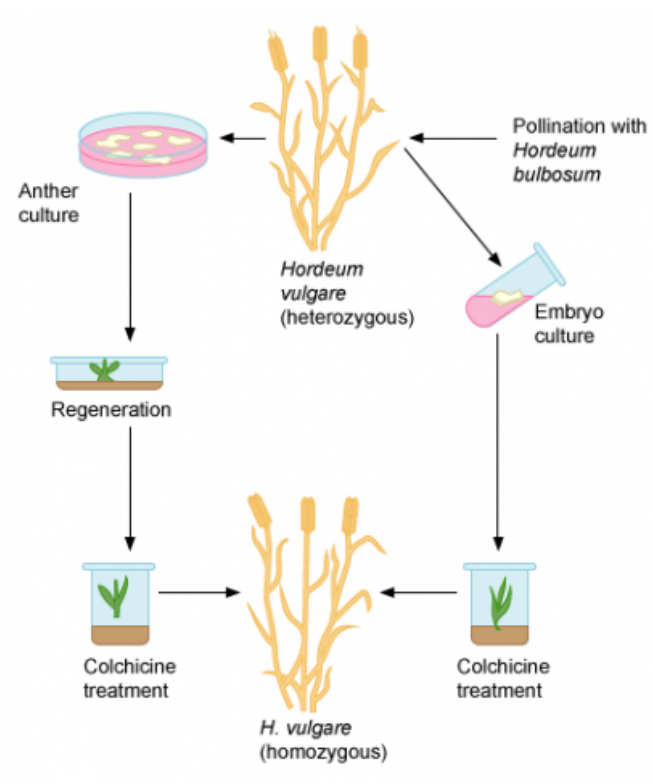


Fig. 5 Development of doubled haploids in barley.

Generally, haploids can be:

- maternal (based on the egg cell)
- paternal (based on microspores / pollen)

Moreover, haploids can be induced by using:

- Tissue culture techniques, such as anther or pollen culture.
- In vivo pollination techniques.

In maize, inducer genotypes are available for both paternal and maternal haploids. Current maternal inducer genotypes, when used as pollinator, result in 10% haploid offspring on the female parent. Respective inducer genotypes are available in other crops such as barley and potato. For some crops, other species are used as inducer. In case of wheat, for example, maize is used for induction of haploids.

Genetics of Polyploidy

Genetic Concepts

Polyploid genetics—especially autoploid genetics—is often more complicated than diploid genetics because genetic ratios are more complex. The inheritance of quantitative characters in polyploids is even more complicated. Before examining the complexity of polyploid genetics, let's become familiar with terminology and associated principles.

- **homoeologous chromosomes** = corresponding chromosomes originating from different but similar genomes. Do not confuse homoeologous chromosomes with homologous chromosomes.

Feature	Homoeologous	Homologous
Synapse of pair during the first division of meiosis?	Sometimes, but only at homologous or partially homologous regions	Yes
Have corresponding loci?	Sometimes	Yes
Derived from same genome?	No	Yes

Valency

- **Valency** = the number of homologous or partially homologous chromosomes that associate or pair during meiosis. A prefix is added to indicate the number of chromosomes that associate. Here are some examples.

Table 5 Valency levels and corresponding number of meiotically-paired chromosomes.

Valency	Number of Meiotically Paired Chromosomes
Univalent	Single
Bivalent	2
Trivalent	3
Quadrivalent	4
Pentavalent	5
Hexavalent	6

Polyploid genotypes can be defined by the

- number of dominant alleles at a locus, or
- number of different alleles at a locus.

Chromosomes

In each of the following examples, note the valency during meiosis and the chromosome constitution of the gametes. Think about what happens when these gametes fuse with normal or other abnormal gametes.

Example 1: Diploid

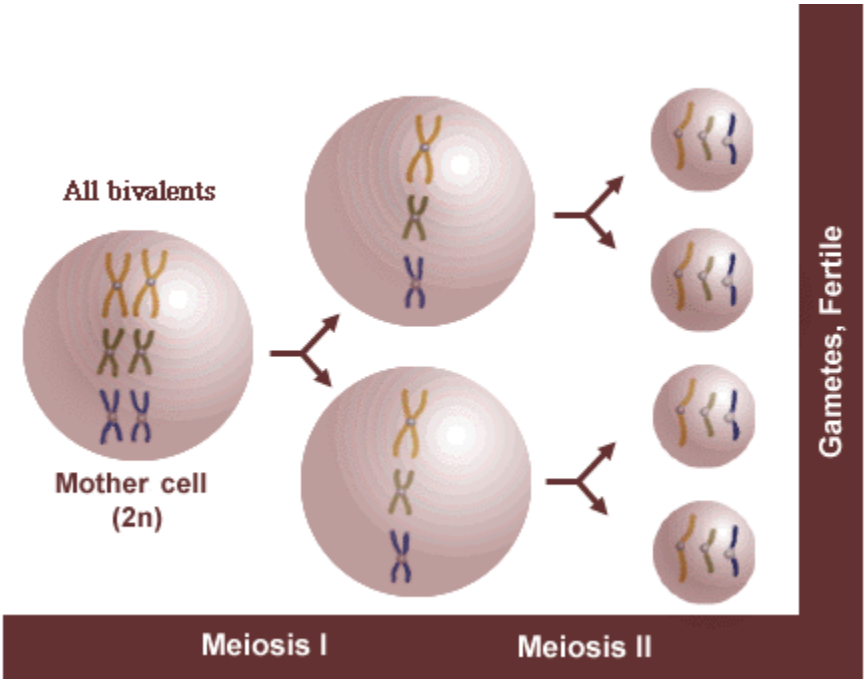


Fig. 6 Source: NIH-NHGRI

If meiosis proceeds normally, each resulting gamete will receive a full complement of chromosomes and be fertile. In this diploid, assume x = 3, n = 3, and 2n = 2x = 6.

Chromosomes form bivalents during meiosis I. One chromosome of each homologous pair migrates to a pole during anaphase I, resulting in a complete chromosome complement in each of two nuclei at the beginning of meiosis II. At anaphase II, sister chromatids separate resulting in a complete chromosome complement in each gamete—gametes are fertile.

Example 2: Autotetraploid

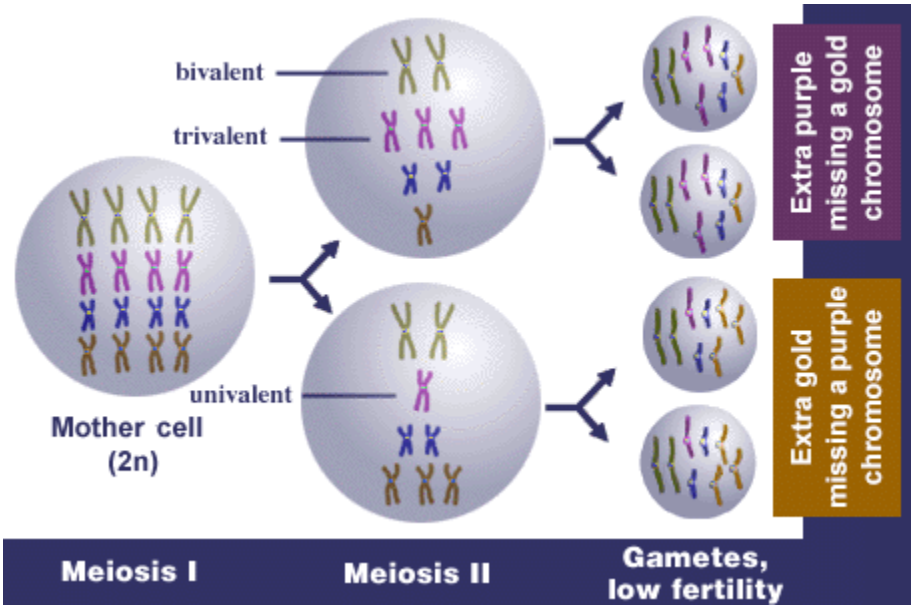


Fig. 7 Source: NIH-NHGRI

Normal synapsis in autotetraploids is as quadrivalents, e.g., homologues from all four genomes synapse. Normal first division separation is that bivalents go to each pole, and thus each gamete contains 2x. Random associations of bivalents can also occur. If homologous or partially homologous chromosomes fail to pair properly or fail to disjoin properly during meiosis, there is more likely to be unbalanced distribution of chromosomes to the daughter cells—some gametes will receive an extra chromosome and others will lack a chromosome, usually resulting in non-viable gametes. Thus, fertility is reduced. Reduced fertility is more common in autoploids than alloploids. Inheritance in autoploids is known as **polysomic inheritance** (or **tetrasomic inheritance** specifically in autotetraploids) and does not follow typical Mendelian patterns.

In this autotetraploid, let x = 4, n = 8, and 2n = 4x = 16.

Example 3: Allotetraploid.

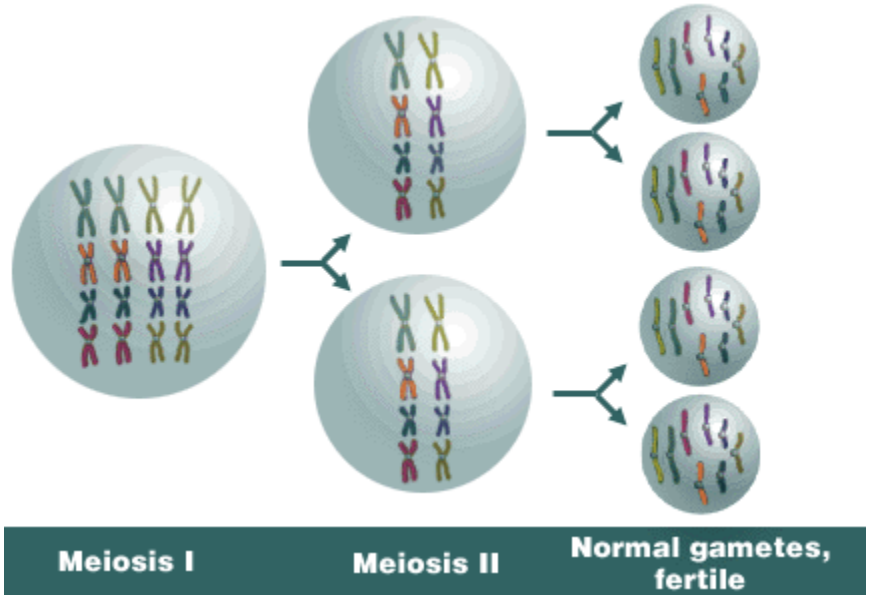


Fig. 8 Source: NIH-NHGRI

Alloploids generally have greater fertility than do autoploids. The formation of bivalents during meiosis contributes to greater fertility. Assume this allotetraploid has x = 4, n = 8, and 2n = 4x = 16.

Alloploid genetics is complex, too, but usually less so than in autoploids. The formation of bivalents in alloploids contributes to their less complex genetics. In addition, bivalency tends to result in disomic inheritance patterns, where two alleles segregate at a locus. Such inheritance of each individual duplicated loci follows typical Mendelian patterns. Thus, the genetic ratios of alloploids more closely resemble those commonly observed in diploids and it's the reason why alloploids are often called amphiploids.

Alloploids that are derived from divergent progenitors exhibit disomic inheritance and are sometimes called genomic alloploids. Segmental alloploids are derived from partially divergent progenitors and typically exhibit mixed inheritance, consisting of both disomic and polysomic inheritance.

Number of Dominant Alleles

The terms used to describe polyploid genotypes encode much information. Let's define

genotypes by the number of times a dominant (**A**) or recessive (**a**) allele is present at a particular locus. The most basic genotype is a **nulliplex**. A nulliplex is a polyploid in which all chromosomes of one homologous type carry the same recessive allele for the particular gene.

Table 6 Nulliplex genotypes for a single locus at different ploidy levels.

Ploidy Level	**Genotype**
Diploid	aa
Triploid	aaa
Tetraploid	aaaa
...	
Octaploid	aaaaaaaa

Other genotypes have at least one dominant allele. The number of dominant alleles is indicated by the term's prefix. Let's look at one locus of an autotetraploid as an example. There are five possible genotypes.

Table 7 Genotypes for a single locus of an autotetraploid.

Term	**Genotype**
Nulliplex	aaaa
Simplex	Aaaa
Duplex	AAaa
Triplex	AAAa
Quadruplex	AAAA

If there is complete dominance, then all of the genotypes except the nulliplex will have the dominant phenotype. Only the nulliplex will exhibit the recessive phenotype. The presence of a dominant allele(s) can mask the presence of recessive alleles. The low frequency of recessive phenotypes necessitates the screening of larger populations in order to detect the presence of recessive alleles.

If there is partial dominance, the phenotypic ratios are very complex. Genetic linkages are extremely difficult to determine. It may even be difficult to identify autotetraploids as such because of intra- and interallelic interactions.

Number of Different Alleles

Let's again use an autotetraploid and consider all the possible genotypes when there are four different alleles at a particular locus. We'll designate these alleles as $\mathbf{a_1}$, $\mathbf{a_2}$, $\mathbf{a_3}$, and $\mathbf{a_4}$ (in this case, however, the use of small letters does not indicate recessive). An individual with all four different alleles present is called a "tetragenic." Again, the terminology tells us much about the genetic composition.

Table 8 Autotetraploid genotypes with four possible alleles at one locus.

Term	Examples of Genotypes	Description
Nulliplex	$a_1a_1a_1a_1$, $a_2a_2a_2a_2$, $a_3a_3a_3a_3$, $a_4a_4a_4a_4$	Only one allele present
Simplex	$a_1a_1a_1a_2$, $a_1a_2a_2a_2$, $a_3a_1a_1a_1$, $a_1a_4a_4a_4$, etc.	One allele present in three copies with another allele present in one copy
Duplex	$a_1a_1a_2a_2$, $a_1a_1a_3a_3$, $a_2a_2a_4a_4$, etc.	Two alleles, each present in two copies
Trigenic	$a_1a_1a_2a_3$, $a_2a_2a_3a_4$, $a_1a_2a_2a_4$, etc.	One allele present in two copies and two different alleles present in one each copy
Tetragenic	$a_1a_2a_3a_4$	Four different alleles present

Autoploid Genetics

Now let's take a closer look at why autoploid genetics is more complex than diploid genetics.

Multiple alleles, partial dominance, additive effects, and epistasis make the inheritance in autoploids more complex and linkage relationships "impossible" to decipher. Multi-allelism also affects the response to breeding procedure.

FURTHER THOUGHT

Many species of crop plants, as well as non-crop plants, are polyploids. Discuss the effects of polyploidy on plants and why this phenomenon is so prevalent.

Autotetraploid Genetics

Chromosomes assort independently in polyploids, just as they do in diploids. In addition, crossing-over can occur between homologous or partially homologous chromosomes. The higher the ploidy level, the more complex the genetics—the greater the number of chromosomes, the greater the number of possible combinations and genotypes that can be produced.

In an autoploid, the number of possible alleles is equal to the ploidy level. Thus, the higher the ploidy level, the greater the possible number of alleles for a given locus. Homologous chromosomes can each carry a different allele at a given locus. Recall that chromatids are the two strands comprising a duplicated chromosome that is still joined at the centromere. When the centromere divides—during anaphase of mitosis or anaphase II of meiosis—then **chromatids** separate and then are called chromosomes. Since chromosomes and chromatids assort randomly, the inheritance patterns are complicated.

Let's examine the case of an autotetraploid. In this case, assume

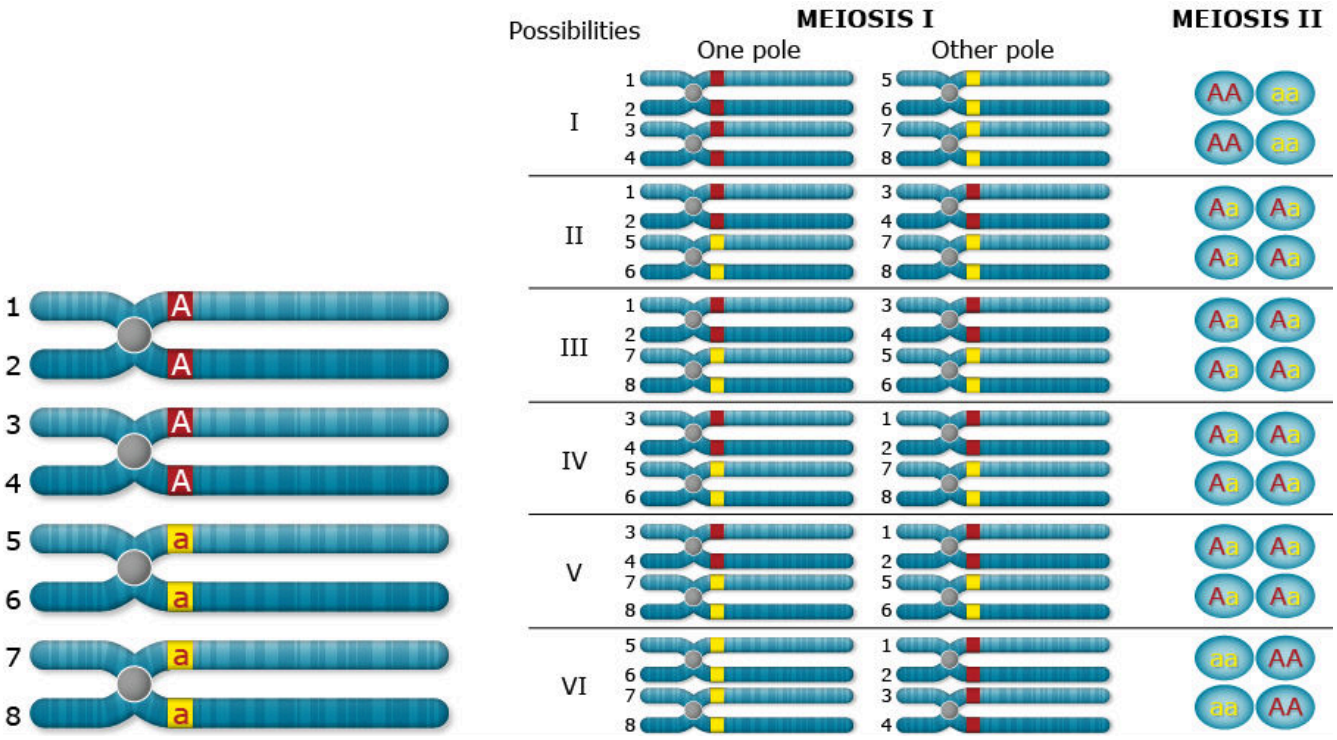
- two alleles at the locus of interest: the "**A**" allele confers complete dominance and the "**a**" allele is recessive; and
- a duplex (**AAaa**) genotype.

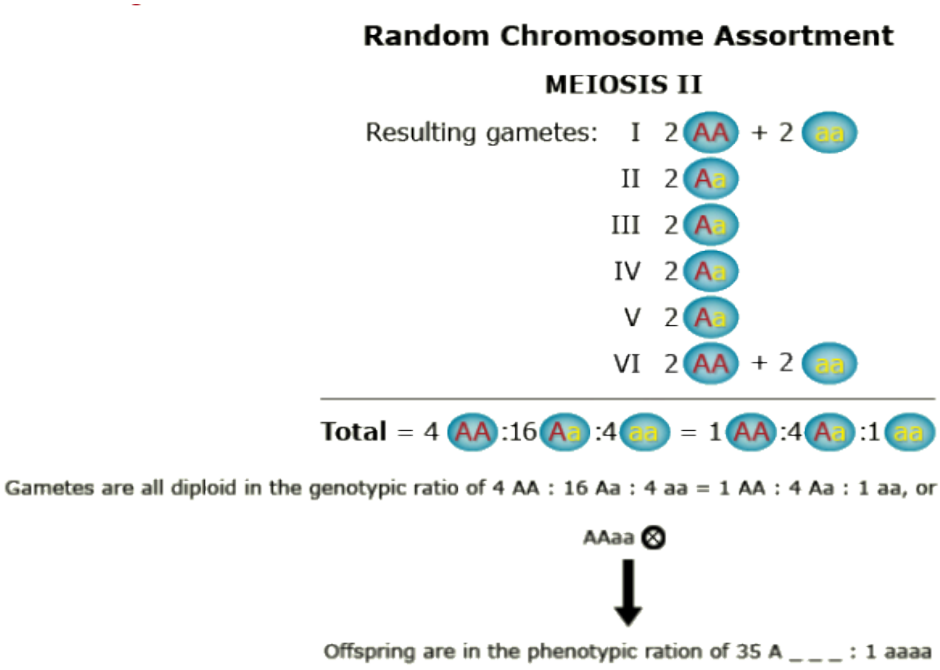
Plants are self-pollinated and the segregation ratios of the progeny can be determined.

Follow the assortment of chromosomes through meiosis to determine the allelic composition of the resulting gametes.

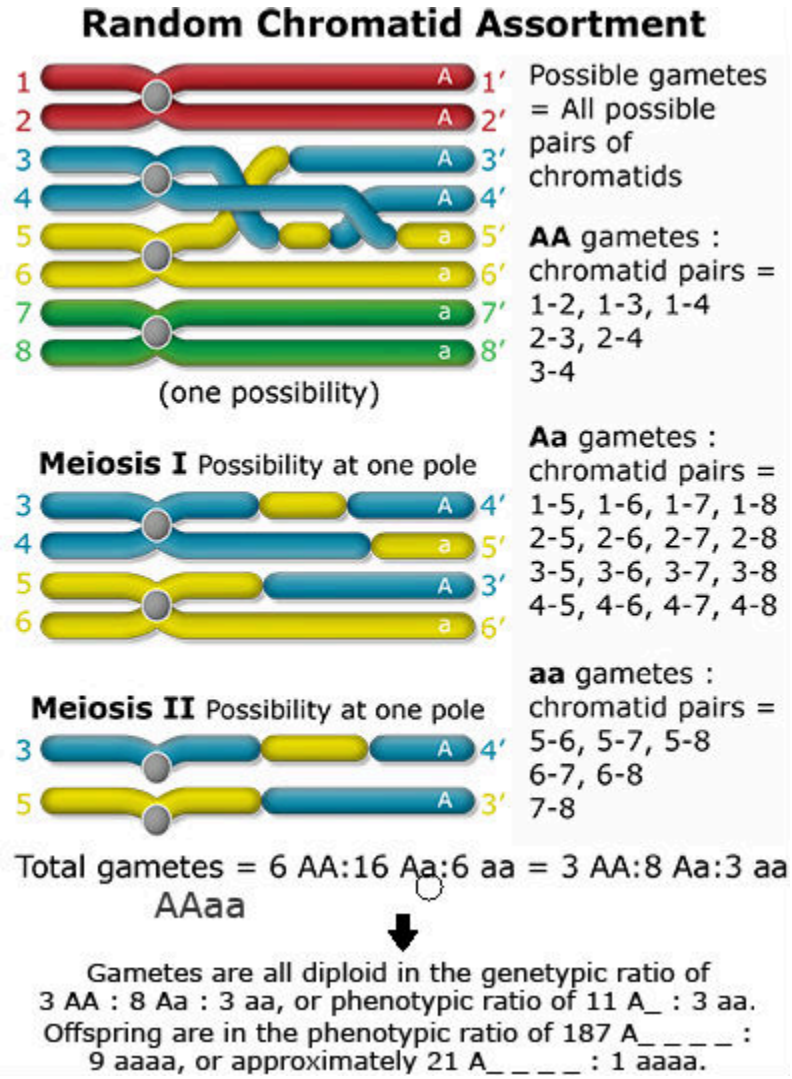
Random Chromosome Assortment

Chromosome inheritance; two alleles, dominance, duplex (**AAaa**)





What happens when recombination occurs between the gene and the centromere? Let's follow the random assortment of the chromatids through meiosis and find out the type of gametes produced.



Conclusions

1. Genetic segregation is more complex in autopolyploids compared to diploids. Homozygous recessive genotypes are more rare in autopolyploids than in diploids. Thus, if breeders intend to fix a particular allele, larger populations are required.
2. There are different models describing genetic segregation in autopolyploids. Two extreme models have been mentioned above: chromosome and chromatid segregation. As shown for the duplex case **AAaa**, both models lead to different segregation rates, with homozygous recessives occurring at frequencies of 1/36 (chromosome segregation) and 1/22 (chromatid segregation). Which of these models is more accurate for a given gene, depends on the distance between this gene and the centromere. The larger the genetic distance, the more likely chromatid segregate independently.

Plant Cell and Tissue Culture

Plant cell and tissue culture methods enable plant breeders to obtain whole plants from somatic or haploid cells or tissues without sexual reproduction. Such methods are beneficial for handling plant material with variation in ploidy that may be otherwise difficult to reproduce. These methods take advantage of the **totipotency** of cells to generate plants.

Under special conditions, cells and tissues extracted from a donor plant can be induced to produce undifferentiated, unorganized cells that are totipotent. Totipotent cells possess all of the genetic information and capabilities needed to produce a whole organism. Because undifferentiated cells have not yet embarked on a developmental track to become a specific tissue or organ, their fate remains undetermined. These totipotent cells can be propagated ***in vitro*** and coaxed into developing roots and shoots, ultimately forming a whole plant.

Tissue culturing has several attractive features:

1. Large numbers of totipotent cells can be cultured in a small area.
2. The culture medium is usually well-defined—each component in the medium is known. Additions or exclusions of specific substances can provide a means of selection for some phenotypes.
3. Ploidy level can be controlled by selection of **explants**.
4. Heritable variation can be induced to generate new genotypes.
5. Tissure culture is often required for genetic transformation.

General Tissue Culture Procedures

Cell and tissue culture procedures are similar. The particulars of this general procedure vary with the cell or tissue source, its age and health, species, genotype, culture type, and culturing objectives.

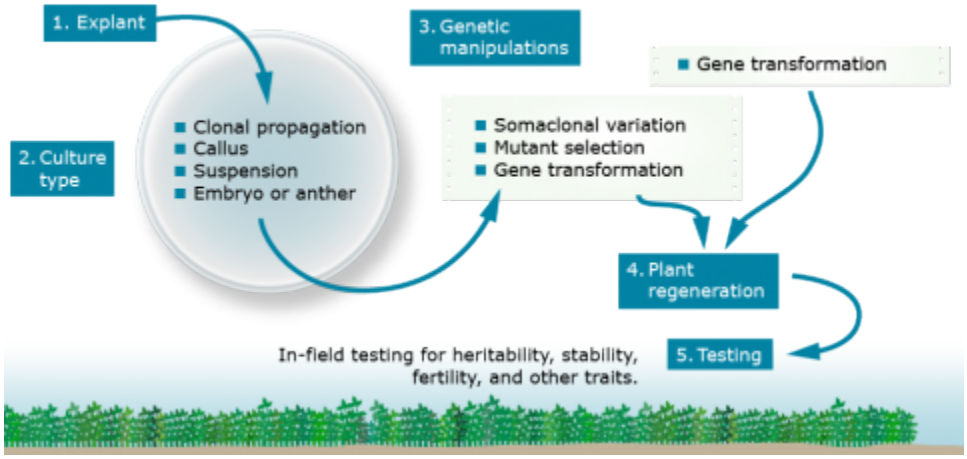


Fig. 9 Cell and tissue culture methods

Types of Tissue Culture

Cell and tissue culture can be used to maintain genotypes or to obtain novel genotypes. Different tissues or types of culture are used to achieve a variety of objectives.

- Clonal propagation or **Micropropagation**—The general purpose of clonal propagation is to reproduce genetically identical plants without going through a gametophytic or sexual stage. A clone is a population of plants derived from a single plant, or a population of cells descendant from a single cell. This is very important in the production of high-value horticultural crops.
- Callus culture—**Callus** culture is similar to clonal propagation, except that the cultured tissue forms disorganized, undifferentiated masses of cells. Spontaneous variants and mutants can occur during callus culturing. Callus culture is often required for successful genetic transformation.
- Suspension culture—Similar to callus culture, except that cells or cell aggregates are in a liquid medium, rather than plated on an agar-solidified medium. These cultures have greater chromosome instability and reduced regeneration potential due to their rapid growth rates. The liquid medium provides relatively uniform exposure of cells to any selecting agents.
- Anther Culture—Anthers containing immature pollen are cultured to develop haploid plants or, with chromosome doubling, to obtain homozygous diploids. Anther culturing involves placing anthers containing immature pollen on medium. Generally, anthers are obtained from F_1 or F_2 plants to maximize the genetic diversity among the haploids

generated through culturing. The greater the genetic diversity, the greater the likelihood of obtaining a desirable genotype to use as breeding lines.

- Embryo culture—Embryo culture allows the rescue of embryos that would normally be unobtainable or aborted because of some type of incompatibility. From the cultured embryos, whole plants may be regenerated.

Embryo Culture

Embryo culture is commonly used to bypass genomic incompatibility. Unfertilized embryos can be used to generate haploid plants or they can be pollinated in vitro to bypass self-incompatibility.

Genomic Incompatibility

Often, embryos resulting from wide crosses (crosses between distantly related parents) are incompatible with the endosperm—the endosperm cannot adequately supply the needs of the growing embryo. As a result of this incompatibility, the embryo will normally abort before the seed is mature. Embryo culture provides a method to circumvent abortion and obtain such hybrids.

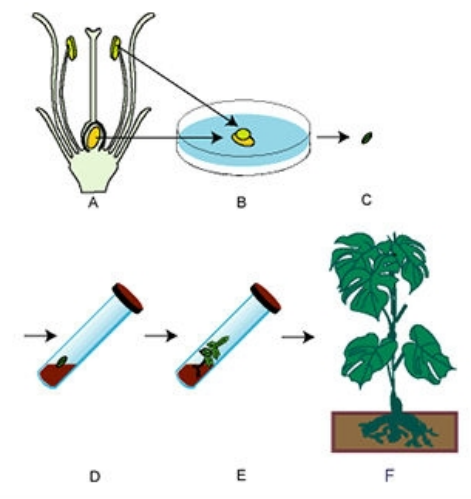


Fig. 10 Embryo culturing of self-incompatible system.

After fertilization, embryos are excised, with or without the ovule, and grown on a nutrient medium. The nutrient medium substitutes for the endosperm in nourishing the growing embryo. Via culturing, hybrid plants can be generated and grown to maturity. These plants will be sterile and must be treated with colchicine to get chromosome doubling. Flowers resulting from the doubled chromosome number will produce viable gametes much like allopolyploids.

Ovule Culture

Unfertilized ovules can be excised and cultured to obtain haploid plants. Alternatively, homozygous diploids can be generated by applying colchicine or other treatments to these

cultured ovules to induce chromosomal doubling. This process is similar to anther culturing and will be discussed more fully in the next section. Ovule culturing has been used with barley, wheat, rice, and maize.

In vitro pollination and fertilization

In self-incompatibility systems where pollen tubes fail to grow, in vitro pollination offers an alternative to overcome such incompatibility. An unfertilized ovule is extracted from the pistillate flower and placed on medium. In vitro pollination and fertilization are accomplished by applying pollen directly to the ovule. Thus, the incompatibility factors localized on the stigma or in the style are avoided. The resulting embryo is nourished by the medium to obtain plantlets.

Uses of Polyploidy, Aneuploidy, and Haploidy

Because of the genetic complexities, breeding polyploids, especially autoploids, is challenging and progress is often slower than breeding diploids. Generally larger populations must be evaluated to obtain favorable types and to stabilize the genotype for the desired phenotype. Breeders also must be concerned about and test for the optimum ploidy level for the particular species and desired characters. Although polyploids present breeding challenges, they also offer opportunities for crop improvement.

ALLOPLOID USES

Alloploids occur more frequently in nature than other types of polyploids. Alloploids can combine the best characters of different species.

Table 9 Examples of alloploid uses

Alloploid Uses	Example
1. Identify genetic origin of crops	Wheat evolved through a combination of three related species. Cultivated *Brassica* species evolved through a series of interspecific hybridizations.
2. Generate new plant genotypes and species	Triticale resulted from a cross of wheat and rye.
3. Enable introgression of genes from related species	Cotton lint strength was improved through transfer of genes from undomesticated cotton into cultivated cotton. (see Fehr, 1987, p. 82 "Introgression of genes")

Generally, alloploids are more vigorous and fertile than autoploids or aneuploids.

AUTOPLOID USES

Autoploids are not as common in nature as alloploids, but in some cases, they are developed as a result of artificially doubling the chromosome number of a diploid species. Depending on the crop, the reduced fertility that tends to occur as a result of induced polyploidy can be an advantage or a disadvantage. Reduced fertility found in autoploids is not a problem for species where the target organ is not seed—for example in forage crops. Fertility tends to be low and vegetative growth greater in autoploids than in diploids. This is an advantage in crops produced for vegetative parts; examples include several tuber and forage crops. Autoploidy is also used to generate sterile plants in which seeds are undesirable, such as in banana and watermelon.

Autoploids are also used to facilitate interspecific or wide crosses. Such crosses allow traits not otherwise available in the crop's gene pool to be acquired, enhancing the genetic diversity of the crop.

ANEUPLOID USES

Aneuploids carry an extra chromosome(s) or are lacking one or more chromosomes — their genomes are unbalanced. Although this usually results in their reduced vigor, their unbalanced

genomes provide benefits for genetic studies and increasing the genetic diversity of a crop. Diploids have very low viability in the aneuploid condition and, therefore, have limited utility. Aneuploidy is more useful in alloploids. Aneuploids have been used to:

- map genes to specific chromosomes
- identify linkage groups
- substitute or transfer chromosomes between related or unrelated species

With molecular techniques available today aneuploids are seldom used to map genes and identify linkage groups.

Chromosome transfer or substitution is a more efficient method than conventional crossing and subsequent segregation and selection because only the particular chromosome carrying the desired gene(s) is transferred. In conventional breeding approaches, both desired and undesired genes are transferred from the donor to the recipient population; subsequently, the desired and adverse traits must be sorted out, often requiring several generations of breeding and selection. Similar to the effects of wide crosses in alloploids, aneuploidy facilitates chromosome transfer from alien species, enabling the recipient population to acquire genes that are otherwise unavailable within its gene pool. Thus, aneuploids can be used to increase genetic diversity.

HAPLOID USES

Haploid plants are of interest in plant breeding and plant genetics because their genomes can be doubled and homozygous lines can be obtained much faster than by conventional techniques of inbreeding. Genetically homogeneous DH lines that can be quickly obtained make it practicable to establish respective experimental populations for gene mapping.

Discussion

Wenzel proposed the analytic–synthetic breeding scheme in potato (shown below). Discuss the strengths and weaknesses of this scheme.

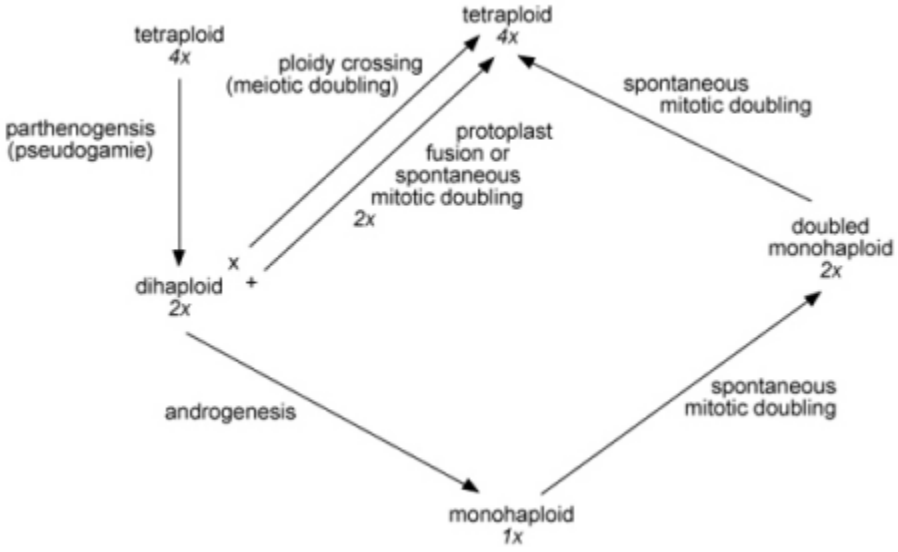


Fig. 11 Wenzel proposed the analytic–synthetic breeding scheme in potato.

References

Becker, Heiko. 1993. Pflanzenzüchtung [Plant Breeding]. Ulmer & Co., Stuttgart, Germany.

Kuckuck, H., G. Kobabe, and G. Wenzel. 1991. Fundamentals of Plant Breeding. Springer Verlag, Berlin, Germany.

Ross, Hans. 1986. Potato breeding – problems and perspectives. In Fortschritte der Pflanzenzchtung [Advances in Plant Breeding]. Verlag Paul Parey, Berlin, Germany.

U, Nagaharu. 1935. Genome analysis in *Brassica* with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. Japan. J. Bot. 7: 389-452.

How to cite this chapter: Muenchrath, D., A. Campbell, L. Merrick, T. Lübberstedt, and S. Fei. (2023). Ploidy: Polyploidy, aneuploidy, haploidy. In W. P. Suza, & K. R. Lamkey (Eds.), *Crop Genetics.* Iowa State University Digital Press. DOI: 10.31274/isudp.2023.130

Contributors

Editors

Walter Suza: Suza is an Adjunct Associate Professor at Iowa State University. He teaches courses on Genetics and Crop Physiology in the Department of Agronomy. In addition to co-developing courses for the ISU Distance MS in Plant Breeding Program, Suza also served as the director of Plant Breeding e-Learning in Africa Program (PBEA) for 8 years. With PBEA, Suza helped provide access to open educational resources on topics related to the genetic improvement of crops. His research is on the metabolism and physiology of plant sterols. Suza holds a Ph.D. in Plant Breeding and Genetics (with emphasis in plant molecular physiology) from the University of Nebraska-Lincoln.

Kendall Lamkey: Lamkey is the chair of the Department of Agronomy at Iowa State University. As chair, he represents the department at the state, national and international levels. He is an advocate for research and the PBEA program, in addition to overseeing the Agronomy Department educational direction, its faculty, as well as Argonomy Extension and Outreach. Dr. Lamkey is a corn breeder and quantitative geneticist and conducts research on the quantitative genetics of selection response, inbreeding depression and heterosis. Lamkey holds a Ph.D. in plant breeding from Iowa State University and a master's in plant breeding from the University of Illinois.

Chapter Authors

William Beavis, Arden Campbell, Jode Edwards, Shui-Zhang Fei, Thomas Lübberstedt, Laura Merrick, and Deborah Muenchrath

Contributors

Gretchen Anderson, Todd Hartnell, Andy Rohrback, Tyler Price, Glenn Wiedenhoeft, and Abbey Elder