SELECTED CHAPTERS OF GENETICS

Lecture Notes for Students of Agricultural Engineering

Dr. habil. Molnár Mónika

Peer reviewed by: Prof. Dr. Sipiczki Mátyás

University of Debrecen

Proofread by: Konczné Nagy Zsuzsanna

University of Nyíregyháza

University of Nyíregyháza, Institute of Environmental Science, 2022

ISBN 978-615-6032-56-0

This project was supported by the Scientific Council of the University of Nyiregyhaza.

Table of contents

1. The Subject and Brief History of Genetics

1.1 What is genetics?

Genetics is a branch of biology. Some define it as a ***"study of heredity"***, though people had been interested in hereditary phenomena long before genetics existed as a discipline. Genetics as a biological discipline began with Gregor Mendel's experiments, who realized the existence of those biological elements that we now call genes.

Genetics is the study of genes. The word genetics comes from the word gene, and genes are in the focus of the subject.

Genetics deals with genes at different research levels:

- reveals the structure of genes;
- examines how genes shape traits in individuals;
- researches the lows according to which they are inherited from parents to offspring;
- examines how the genetic diversity of a population shapes evolution.

Whether geneticists study at the molecular, cellular, organismal, family, population or evolutionary level, genes are always central in their studies.

What is the relationship between the organism and the genes?

An organism's basic complement of DNA is called its ***genome***. The body cells of most plants and animals contain two genomes. These organisms are ***diploid***. The hereditary material is organized into ***chromosomes*** in the ***nuclei*** of cells. Diploid cells contain two sets of chromosomes. For instance, human body cells contain two sets of 23 chromosomes, for a total of 46. There is a ***single DNA strand in each chromosome*** (Figure 1.). ***A gene is the functional region of the chromosomal DNA.***

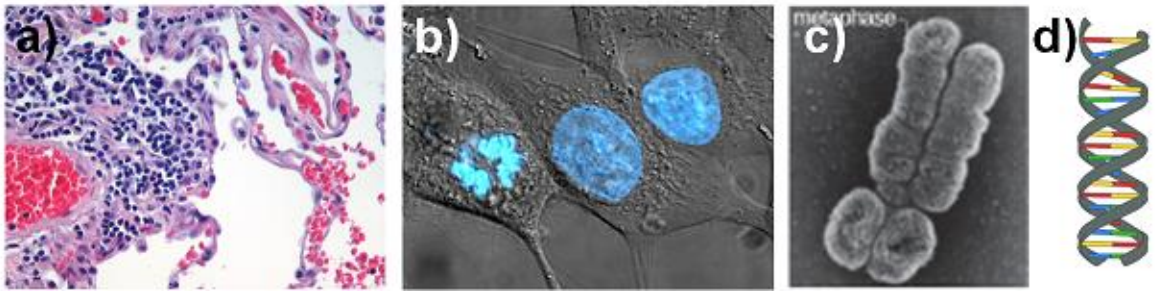


Figure 1. Focusing on the genetic material

a) The body of organisms is built up from different tissues. This histological specimen is e.g. a human lung tissue. b) Tissues are made up of cells; each of them contains a nucleus. The DNA-binding dye stains blue the nuclei of these three cells, the leftmost of which is in division. c) This chromosome in the metaphase stage of cell division consists of two sister chromatids. Each sister chromatid contains a DNA strand. d) Schematic representation of the DNA.

a) https://commons.wikimedia.org/wiki/File:Emphysema_H_and_E.jpg; CC BY 2.0 https://creativecommons.org/licenses/by/2.0/deed.en ;
b) https://commons.wikimedia.org/wiki/File:HeLa_cells_stained_with_Hoechst_33258.jpg; TenOfAllTrades at English Wikipedia, Public domain, via Wikimedia Commons
c) https://commons.wikimedia.org/wiki/File:Stages_of_early_mitosis_in_a_vertebrate_cell_with_micrographs_of_chromatids.svg David O Morgan, Attribution, via Wikimedia Commons
d) https://commons.wikimedia.org/wiki/File:Eukaryote_DNA-en.svg; Sponk, Tryphon, Magnus Manske, User:Dietzel65, LadyofHats (Mariana Ruiz), Radio89, CC BY-SA 3.0 <https://creativecommons.org/licenses/by-sa/3.0>, via Wikimedia Commons

Genes dictate the inherent properties of a species. The products of most genes are specific ***proteins*** that are the main macromolecules of an organism. ***The amino acid sequence of a protein is encoded by a gene.***

How do genes perform their biological role?

Three fundamental properties are required of genes and the DNA of which they are composed:

1. DNA must ***transmit the genetic information faithfully*** from cell to cell, from generation to generation.
2. Through its ***controlled expression***, DNA must create the organism's cells/body.
3. DNA must be able to ***undergo heritable changes***.

1. ***DNA replication*** is the basis of the perpetuation of life. Replication occurs at two stages of the life cycle of an organism (Figure 2.). Once, before gametes (eggs and sperms) are produced by cell division. These cells ensure the continuation of the species from one generation to the next. Then, replication must occur each time when the fertilized egg, that is the zygote, divides repeatedly to produce the body of a new individual.

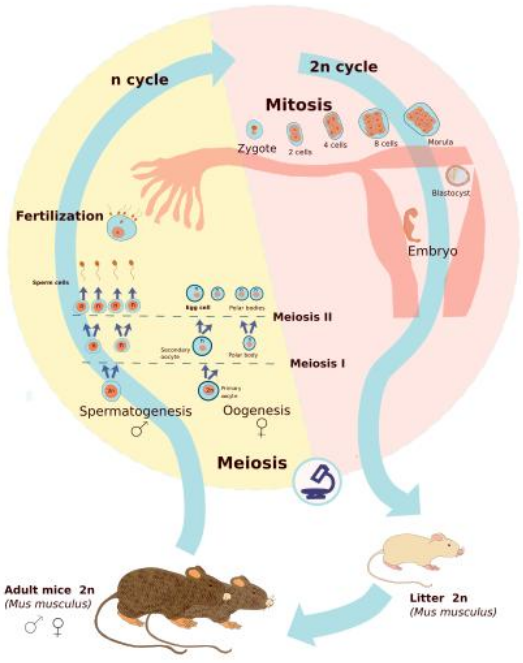


Figure 2. Diploid life cycle of the mammalian model organism, mouse (*Mus musculus*)

DNA replication occurs in the male and female parent before meiotic divisions produce sperms and eggs, during the spermatogenesis and oogenesis, respectively. Then, successive mitotic divisions produce an embryo, and later a baby mouse from the fertilized egg.
https://commons.wikimedia.org/wiki/File:Diploid_English.svg

2. ***Controlled expression*** of the DNA's information generates the form of an organism (Figure 3.)

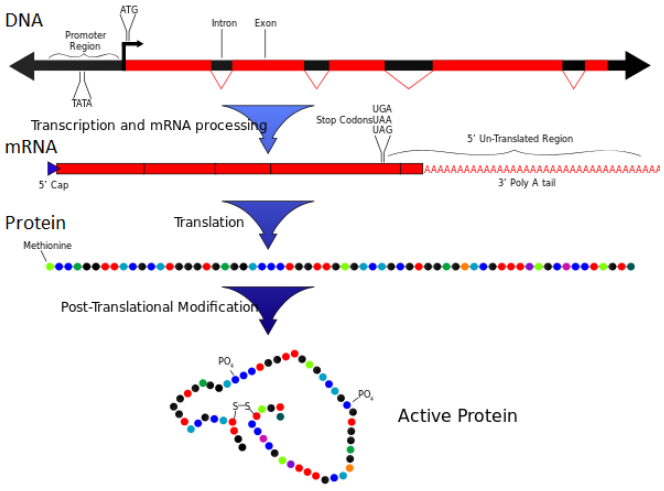


Figure 3. The flow of genetic information in a eukaryotic cell

"Controlled gene expression" means the regulated transcription and translation of the information encoded by the genes. These steps of the information transfer will be discussed in Chapter 3.
https://commons.wikimedia.org/wiki/File:Cdmb.svg
Adenosine at English Wikipedia, CC BY-SA 2.5 <https://creativecommons.org/licenses/by-sa/2.5>, via Wikimedia Commons

The basic flow of genetic information is from DNA to RNA to protein. Traditionally, this statement is called the ***"central dogma"*** of genetics.

3. Genes undergo ***heritable changes*** rarely but regularly. The process when a gene changes from one form to another is called ***mutation*** (Figure 4.). Mutation is not only a basis for variation within a species, but over the long time it provides raw material for the evolution.

Figure 4. *Drosophila* mutants

Four mutant flies with undeveloped wings that carry the "vestigial" mutation, and a normal individual (upper left). Mutants are useful experimental tools for the researchers.
https://commons.wikimedia.org/wiki/File:Drosophila_melanogaster_vestigial.jpg; Salem.slima, CC BY-SA 4.0 <https://creativecommons.org/licenses/by-sa/4.0>, via Wikimedia Commons.
Upper left inset: https://commons.wikimedia.org/wiki/File:Drosophila_melanogaster_-_side_(aka).jpg; André Karwath aka Aka, CC BY-SA 2.5 <https://creativecommons.org/licenses/by-sa/2.5>, via Wikimedia Commons

1.2 Chapters from the history of genetics

Early “geneticists” – before Mendel

Different aspects of inheritance had been of interest to humans long before biology or genetics existed. Ancient people were improving plant crops and domesticated animals by ***selecting desirable individuals for breeding*** (Figure 5.).

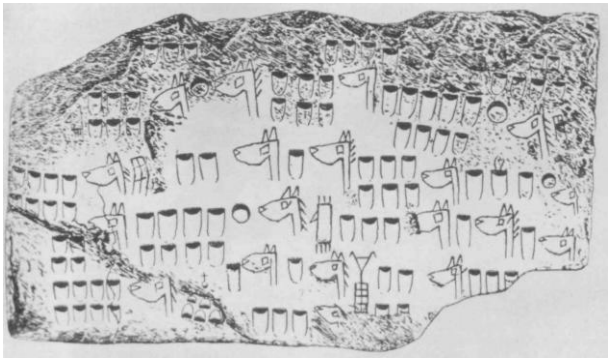


Figure 5. Possibly the oldest pedigree in existence

This engraved seal from an excavation in Mesopotamia is about 5000 years old. Note the three mane types and three profile types.
Image: *Journal of Heredity*, Volume 26, Issue 6, June 1935, Page 237.

The invention of the microscope contributed greatly to the development of cell biology and genetics. Early microscopists started to explore the microbial life and examined different cell types during the 17th century. The English polymath Robert Hooke and the Dutch businessman and scientist Antonie van Leeuwenhoek were the most prominent pioneers in this field (Figure 6.)

Figure 6. Early microscopists and their instruments

Robert Hooke (1635-1703) and his microscope (left); Antonie van Leeuwenhoek (1632-1723) and his microscope (right). Hooke constructed the first microscope and was the first to see “cells”. The small, empty compartments he observed in the cross section of cork reminded him the chambers (cella, Latin) of monks, hence the term “cell”. Leeuwenhoek built the best microscopes of his age and observed several different cell types.
https://commons.wikimedia.org/wiki/File:Portrait_of_a_Mathematician_1680c.jpg; Mary Beale, Public domain, via Wikimedia Commons; https://commons.wikimedia.org/wiki/File:Hooke_Microscope-03000276-FIG-4.jpg; See page for author, Public domain, via Wikimedia Commons;
https://commons.wikimedia.org/wiki/File:Anthonie_van_Leeuwenhoek_(1632-1723)._Natuurkundige_te_Delft_Rijksmuseum_SK-A-957.jpeg; Jan Verkolje, Public domain, via Wikimedia Commons; https://commons.wikimedia.org/wiki/File:Leeuwenhoek_Microscope.png; Jeroen Rouwkema, CC BY-SA 3.0 <https://creativecommons.org/licenses/by-sa/3.0>, via Wikimedia Commons

The observation of cells finally led to the formulation of the ***cell theory (1839)***. Developing this theory is attributed to two German scientists: the animal physiologist Theodore Schwann, and the botanist Matthias Jakob Schleiden (Figure 7.).

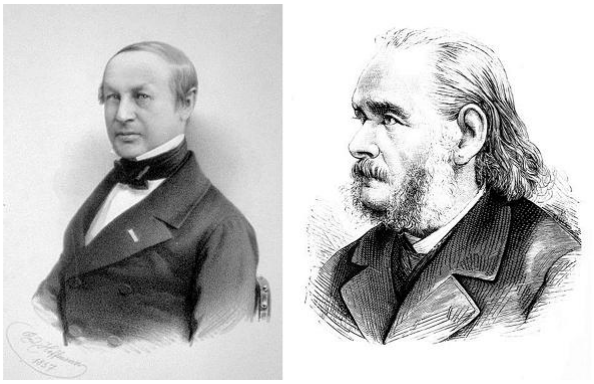


Figure 7. Fathers of the cell theory

Theodor Schwann (1810-1882) left; Matthias Jakob Schleiden (1804-1881) right
https://commons.wikimedia.org/wiki/File:Theodor_Schwann_Litho.jpg; Rudolph Hoffmann, Public domain, via Wikimedia Commons;
https://commons.wikimedia.org/wiki/File:PSM_V22_D156_Matthias_Jacob_Schleiden.jpg; Unknown author, Public domain, via Wikimedia Commons

The three tenets of cell theory:

1. ***All living organisms are composed of one or more cells.***
2. ***The cell is the basic unit of structure and organization in the organisms.***
3. ***Cells arise from pre-existing cells.***

A further step towards the foundation of genetics as a science was the study of cell division (Figure 8.). The ***two types of cell division*** and ***chromosomes*** were first described in the late 19th century.

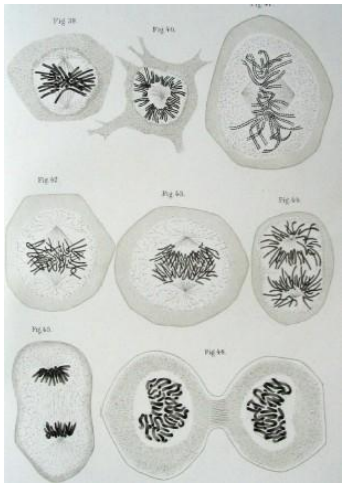


Figure 8. Mitotically dividing cells with chromosomes

Walter Flemming's illustration from 1882. Flemming investigated the distribution of chromosomes into daughter nuclei in mitosis. He is considered as the founder of cytogenetics. Meiotic chromosomes were first described by the Belgien zoologist, Edouard Van Beneden, in 1883.
https://commons.wikimedia.org/wiki/File:Zellsubstanz-Kern-Kerntheilung.jpg
Walther Flemming (1843-1905), Public domain, via Wikimedia Commons

“Classical genetics” – Mendel and beyond

Gregor Mendel: founder of the science of genetics

Gregor Johann Mendel (Figure 9.) was an Augustinian friar and abbot of St Thomas Abbey in Brno. Mendel studied the ***garden pea***. The experiments he carried out in the ***1860s*** revealed the rules of heredity. In modern genetics, these rules are referred to as the ***laws of Mendelian inheritance***. Mendel’s work is an excellent example of how to choose proper experimental methods and devise an experiment correctly; therefore, his experiments are regularly being taught in textbooks. There is a flower bed in the garden of his monastery that reminds us of his experiments.

Figure 9. Gregor Mendel (1822-1884)

https://commons.wikimedia.org/wiki/File:Gregor_Mendel_2.jpg
Unknown author, Public domain, via Wikimedia Commons

The chromosome theory of inheritance (1902)

The chromosome theory of inheritance is credited to the German zoologist and comparative anatomist Theodor Boveri, and the American geneticist and physician Walter Sutton (Figure 10.). This fundamental theory of genetics formulates a connection between the cell division and the rules of Mendelian inheritance. It also states that ***the genes***, which were still unidentified units of inheritance at that time, ***are parts of the chromosomes***. Chromosome segregation studies carried out in insect cells, especially those that investigated the sex chromosomes, helped to prove the Sutton-Boveri theory.

Figure 10. Fathers of the chromosome theory of inheritance

Walter Sutton (1877-1916) left; Theodor Boveri (1862-1915) right
https://commons.wikimedia.org/wiki/File:Walter_sutton.jpg; Unknown author, Public domain, via Wikimedia Commons
https://commons.wikimedia.org/wiki/File:Theodor_Boveri_high_res-2.jpg; Unknown author, Public domain, via Wikimedia Commons

Morgan's Fly Room

Thomas Hunt Morgan (Figure 11.) studied the genetic characteristics of the ***fruit fly Drosophila melanogaster*** at Columbia University, in a laboratory that became famous as the Fly Room. As a result of Morgan and his coworkers' work, the fruit fly became a major model organism in contemporary genetics. ***Morgan proved that genes are carried by chromosomes and are the material bases of heredity.***

Figure 11. Thomas Hunt Morgan (1866-1945)

https://commons.wikimedia.org/wiki/File:Thomas_Hunt_Morgan.jpg
Unknown author, Public domain, via Wikimedia Commons

Many *Drosophila* mutants, especially ***mutants of the eye color and wing***, were isolated and examined (see Figure 4, on page 5). Notable is the white-eyed mutation that Morgan used in the study of sex-linked inheritance. From the accumulated results of recombination studies, Morgan's student Alfred Sturtevant developed the ***first genetic map*** in 1913 (Figure 12.).

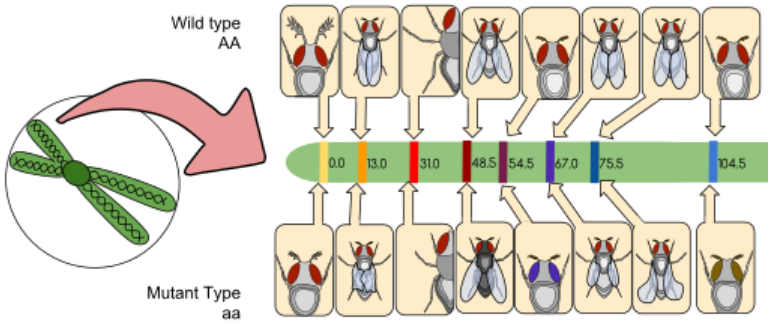


Figure 12. A gene linkage map, based on Morgan's crosses

The relative position of gene loci on chromosome *II* and the phenotype of mutants used in the crosses are presented in this modern drawing.
https://commons.wikimedia.org/wiki/File:Drosophila_Gene_Linkage_Map.svg
Twaanders17, CC BY-SA 4.0 <https://creativecommons.org/licenses/by-sa/4.0>, via Wikimedia Commons

Modern genetics – the age of DNA

Griffith's transforming principle (1928)

Frederick Griffith (Figure 13.) was a British bacteriologist, whose work focused on the epidemology and pathology of pneumonia. By using two strains of the bacterium *Streptococcus pneumoniae*, he demonstrated that ***bacteria are capable of transferring genetic information through a process called transformation***. With his fundamental experiments, the search for the ***transforming principle/agent*** began.

Figure 13. Frederick Griffith (1879-1941)

https://commons.wikimedia.org/wiki/File:Fred_Griffith_and_%22Bobby%22_1936.jpg
Coburn, Alvin F., Public domain, via Wikimedia Commons

One-gene – one-enzyme hypothesis (1941)

The one-gene – one-enzyme hypothesis is credited to two American geneticists, George Beadle and Edward Tatum. Beadle and Tatum experimented with auxotrophic mutants of the bread mold, *Neurospora crassa* (Figure 14.).

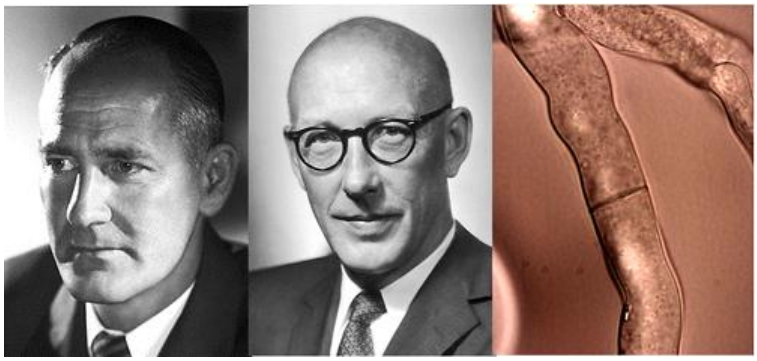


Figure 14. The protagonists of the one-gene – one-enzyme story: two scientists and their experimental organism

George Beadle (1903-1989) left; Edward Tatum (1909-1975) middle; Hyphae of the fungus *Neurospora crassa*
https://en.wikipedia.org/wiki/File:George_Wells_Beadle.jpg#file; Source: http://www.nobelprize.org/nobel_prizes/medicine/laureates/1958/ Author: Unknown.
https://en.wikipedia.org/wiki/File:Edward_Lawrie_Tatum_nobel.jpg; Source: http://www.nobelprize.org/nobel_prizes/medicine/laureates/1958/, Author: Unknown.
https://commons.wikimedia.org/wiki/File:Neurospora_crassahyphae.jpg; Roland Gromes, CC BY-SA 3.0 <https://creativecommons.org/licenses/by-sa/3.0>, via Wikimedia Commons

The experimental results suggested a link between genes and enzymatic reactions, which they formulated as the one-gene – one-enzyme hypothesis. The one-gene – one enzyme hypothesis is one of the great ***unifying concepts*** in biology, because it brings together the concepts and research techniques of genetics and biochemistry. The main statements of this theory can be summarized as follows: 1. Biochemical reactions in vivo consist of a series of discrete, step-by-step reactions. 2. Each reaction is specifically catalyzed by a single enzyme. 3. Each enzyme is specified by a single gene. ***Genes control biochemical reactions by producing enzymes.***

The Hersey-Chase experiment (1952)

The crucial experiment that helped to confirm that DNA is the genetic material was carried out by the American geneticists Alfred Hershey and Martha Chase, with bacteriophage T2 (Figure 15). The scientists used selective radioactive labeling of the DNA and protein of the phage, whereby they could demonstrate that ***the hereditary material*** of the phage ***is DNA***, not protein.

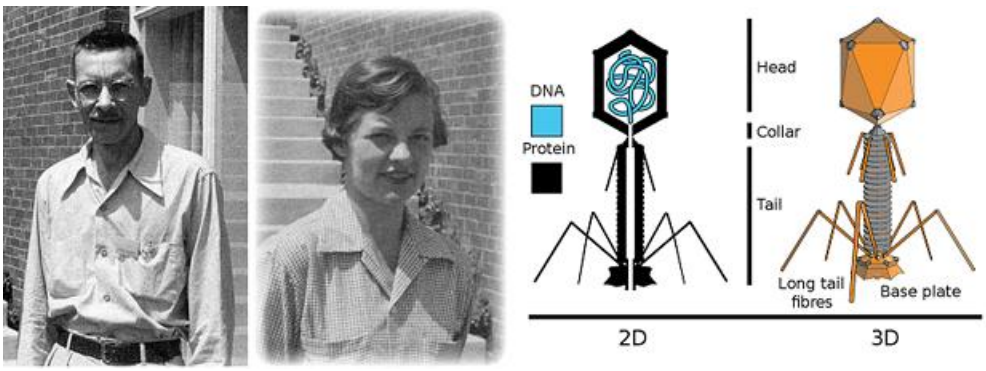


Figure 15. Scientists and their experimental subject

Alfred Hershey (1908-1997), Martha Chase (1927-2003), and a 2D and 3D representation of the structure of a T-even phage. T-even phages have become widely-used experimental systems in genetics.
https://commons.wikimedia.org/wiki/File:Alfred_Hershey.jpg; Unknown author, Public domain, via Wikimedia Commons
https://commons.wikimedia.org/wiki/File:Martha_Chase.jpg; Unknown author, Public domain, via Wikimedia Commons
https://commons.wikimedia.org/wiki/File:Tevenphage.svg; Adenosine (original); en:User:Pbroks13 (redraw), CC BY-SA 2.5 <https://creativecommons.org/licenses/by-sa/2.5>, via Wikimedia Commons

Deciphering the structure of DNA (1953)

The ***molecular structure of the DNA*** was deciphered by the American geneticist James Watson and the English biophysicist Francis Crick. Together ***they built a famous model of the double helix***, to which they used, among others, the results of X-ray diffraction experiments obtained by Maurice Wilkins and Rosalind Franklin (Figure 16.). They published the structure in the prestigious scientific journal Nature, on the 25th April, 1953. This date can be called the birthday of molecular genetics. The first sentences of the article are: "We wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest."

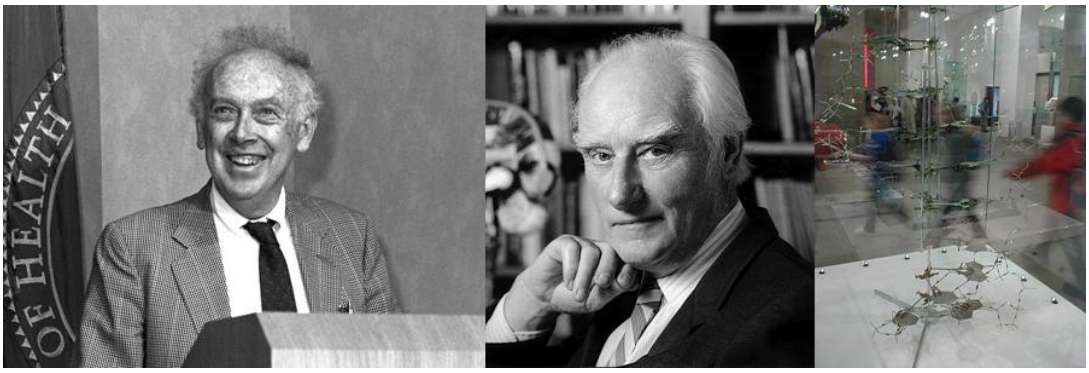


Figure 16. Watson, Crick, and the famous model

James Watson (born 1928) left; Francis Crick (1916-2004) middle. The model built by Crick and Watson is now on display in the National Science Museum of London.
https://commons.wikimedia.org/wiki/File:James_Dewey_Watson.jpg; National Cancer Institute (NCI), Public domain, via Wikimedia Commons; https://commons.wikimedia.org/wiki/File:Francis_Crick_crop.jpg; Francis_Crick.png: Photo: Marc Liebermanderivative work: Materialscientist, CC BY 2.5 <https://creativecommons.org/licenses/by/2.5>, via Wikimedia Commons;
https://commons.wikimedia.org/wiki/File:DNA_Model_Crick-Watson.jpg; User:Alkivar, Public domain, via Wikimedia Commons

Deciphering the genetic code (1961)

The research groups that undertook to decipher the genetic code were led by the American biochemist and geneticist Marshall Warren Nirenberg, and the Indian American biochemist Har Gobind Khorana (Figure 17.) Efforts to understand ***how proteins are encoded by DNA*** began right after the structure of DNA was discovered. Scientists used a cell-free *E. coli* system and synthetic mRNAs. The first result: the codon UUU specifies the amino acid phenylalanine.

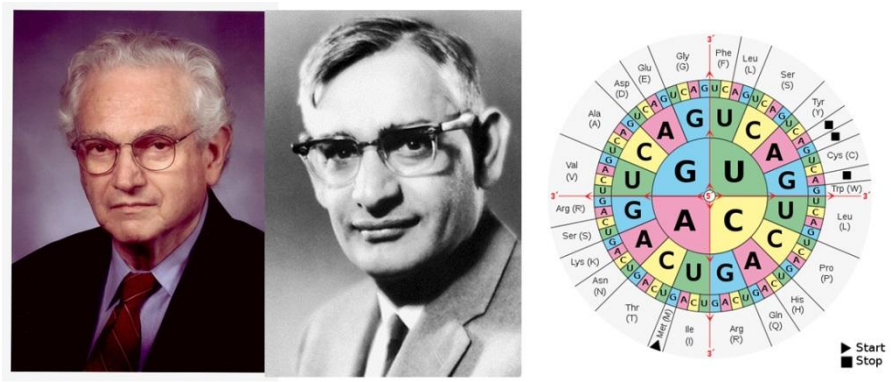


Figure 17. Nirenberg, Khorana, and the genetic code

Marshall Nirenberg (1927-2010) left; Gobind Khorana (1922-2011) middle; and one of the many colorful presentations of the system of genetic code.
https://commons.wikimedia.org/wiki/File:Marshall_Nirenberg_(2002,_by_National_Institutes_of_Health,_U.S.).jpg; National Library of Medicine, No restrictions, via Wikimedia Commons
https://en.wikipedia.org/wiki/File:Har_Gobind_Khorana_nobel.jpg
Source: http://www.nobelprize.org/nobel_prizes/medicine/laureates/1968/ Author: Unknown
https://commons.wikimedia.org/wiki/File:Codons_aminoacids_table.png; Mouagip, Public domain, via Wikimedia Commons

Modern genetics – the age of genomics

After solving the structure of DNA and deciphering the genetic code, researchers in the 1970s and 1980s invented several new methods that started a new era in genetics. These methods have produced enormous amount of information about the genetic model systems and changed the way how genetic problems are approached fundamentally. Nowadays, in the age of genomics, the complete genome sequence of thousands of species is stored in databases, opening the way to their analysis and genetic manipulation. Next, a few basic methods of molecular genetics and the pioneering genome programs will be mentioned briefly.

Sequencing (1977)

In genetics and biochemistry, sequencing means the determination of the primary structure of a biological macromolecule, DNA, RNA or protein. In practice, when sequencing is mentioned we think of the ***determination of the order of the building blocks (nucleotides) of DNA.*** Initially two methods: a chemical method developed by Maxam and Gilbert, and the dideoxy-sequencing worked out by the English biochemist, Frederick Sanger (Figure 18.), were applied. Later the Sanger sequencing has become prevailing, and after automatization almost exclusive. So far, most DNA sequencing has been performed using Frederick Sanger's method and its improved versions. ***The first complete genome ever sequenced*** was that of the ***bacteriophage PhiX174*** or ΦX174.

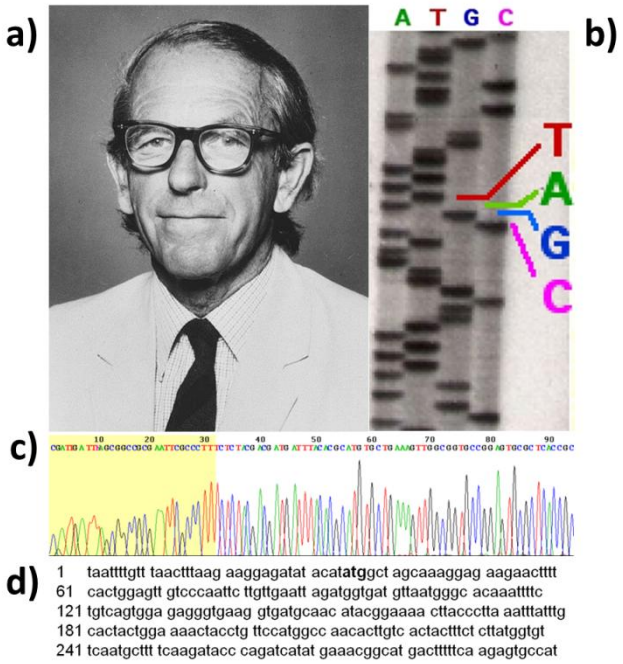


Figure 18. Sequencing

a) Frederick Sanger (1918-2013), inventor of the dideoxy-, also called chain terminator sequencing. b) Part of a radioactively labelled sequencing gel. c) Result of an automated sequencing that used fluorescent dyes and dye-terminator method. d) Nucleotide sequence encoding a portion of the GFP retrieved from GenBank.
https://commons.wikimedia.org/wiki/File:Frederick_Sanger2.jpg
source: https://www.nlm.nih.gov/visibleproofs/media/gallery/vi_a_208b.jpg
https://commons.wikimedia.org/wiki/File:Sequencing.jpg;
John Schmidt, CC BY-SA 3.0 <http://creativecommons.org/licenses/by-sa/3.0/>, via Wikimedia Commons
https://en.wikipedia.org/wiki/File:Sanger_sequencing_read_display.gif#file
Loris, Public domain, via Wikimedia Commons

Polymerase chain reaction (PCR) (1983)

PCR is a method of in vitro cloning widely used to rapidly ***amplify a specific DNA sample.*** PCR was invented by the American biochemist Karry Mullis (Figure 19.), in 1983. Two major difficulties had to be overcome when developing the method. A machine was needed that could carry out quick temperature changes during the thermal cycling, and a heat-stable DNA polymerase that could resist the high temperature used for DNA denaturation. Since then, PCR has become fundamental to many of the procedures used in genetic testing and research. Potential applications of the technique include for example, DNA cloning for sequencing, gene mutagenesis, diagnosis and monitoring of genetic disorders, making genetic fingerprints, detection of pathogens, etc.

Figure 19. Karry Mullis and a PCR machine

Karry Mullis (born 1944), the inventor of PCR, and one of the many types of PCR equipment. A researcher is placing a strip of eight PCR tubes into the thermal cycler.
https://commons.wikimedia.org/wiki/File:Kary_Mullis.jpg
Dona Mapston, CC BY-SA 3.0 <https://creativecommons.org/licenses/by-sa/3.0>, via Wikimedia Commons
https://commons.wikimedia.org/wiki/File:PCR_masina_kasutamine.jpg
Karl Mumm, CC BY-SA 3.0 <https://creativecommons.org/licenses/by-sa/3.0>, via Wikimedia Commons

The most remarkable genome programs

1995: The genome of the ***bacterium Haemophilus influenzae*** is the ***first genome of a free-living organism*** to be sequenced. Interest in *H. influenzae* was raised by the fact that it was incorrectly described as the causative microbe of flu.

1996: ***Saccharomyces cerevisiae, a yeast species, is the first eukaryote*** whose genome was ever sequenced. Budding yeast is an excellent model organism for the investigation of several biological processes, like e.g. cell cycle, meiosis, signal transduction, mitochondrial disorders, etc.

1998: The first ***genome sequence of a multicellular eukaryote, Caenorhabditis elegans***, is released. *C. elegans*, the free-living transparent nematode, has also been used as a model organism (Figure 20.).

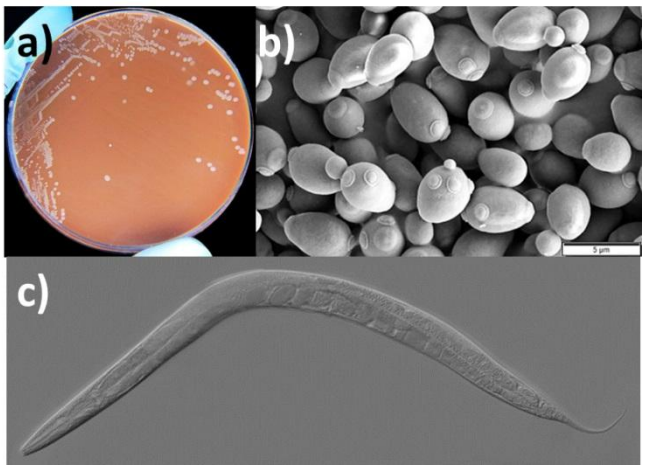


Figure 20. “Pioneer organisms” in genome sequencing

a) *H. influenzae* on an agar plate. b) *S. cerevisiae*, SEM image. c) An adult hermaphrodite *C. elegans*
https://commons.wikimedia.org/wiki/File:Haemophilus_influenzae.jpg
Stefan Walkowski, CC BY-SA 4.0 <https://creativecommons.org/licenses/by-sa/4.0>, via Wikimedia Commons
https://commons.wikimedia.org/wiki/File:Saccharomyces_cerevisiae_SEM.jpg; Mogana Das Murtey and Patchamuthu Ramasamy, CC BY-SA 3.0 <https://creativecommons.org/licenses/by-sa/3.0>, via Wikimedia Commons
https://commons.wikimedia.org/wiki/File:Adult_Caenorhabditis_elegans.jpg; The original uploader was Kbradnam at English Wikipedia.(Original text: Zeynep F. Altun, Editor of www.wormatlas.org), CC BY-SA 2.5 <https://creativecommons.org/licenses/by-sa/2.5>, via Wikimedia Commons

The Human Genome Project (HGP)

The HGP was an international scientific enterprise launched in 1990 with the goal of sequencing the whole human genome. Most of the work was carried out at universities and in research centers in the United States, the United Kingdom, Japan, France, Germany, and China. Researchers in the international consortium used a mixed sample from anonymous donors. A parallel project with the same goal was also launched by the Celera Corporation or Celera Genomics, in 1998 (Figure 21.). The competition of the “opponents” has greatly contributed to the rapid development of the sequencing methods. The first draft sequences of the human genome were released simultaneously by the HPG and Celera Genomics in 2001. ***Successful completion of the sequencing of the human genome was announced in 2003***.

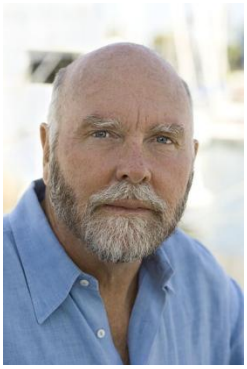


Figure 21. Craig Venter, founder of the Celera Genomics

Craig Venter is the first man whose personal genome sequence was determined.
https://commons.wikimedia.org/wiki/File:Craigventer2.jpg; Article by Liza Gross, but no photo credit given, CC BY 2.5 <https://creativecommons.org/licenses/by/2.5>, via Wikimedia Commons

2. The Organization of the Hereditary Material: Cytogenetic Basics

2.1 The prokaryotic and the eukaryotic cell
(Cell biology reminder)

All living organisms are composed of cells. Cells are of two types: prokaryotic and eukaryotic. ***Prokaryotic cells*** (Figure 22.) are smaller, simpler and more ancient than eukaryotic cells. Prokaryotes are unicellular organisms that can be divided into two domains of life: ***Bacteria*** and ***Archaea***. Prokaryotes lack a membrane enclosed nucleus (pro = "before"; karyon = "nucleus"), and do not possess membrane-enclosed organelles. The prokaryotic hereditary material is a ***single, circular DNA*** which is localized in the ***nucleoid region*** of the cell. In the prokaryotic "chromosome", the DNA is not associated with special proteins.

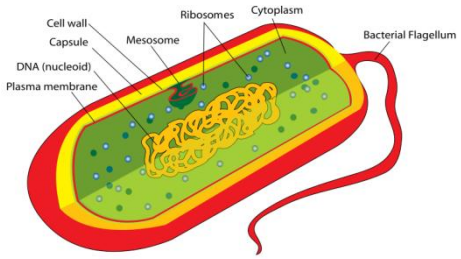


Figure 22. Prokaryotic cell

The plasma membrane, cytosol, ribosomes, and the nucleoid region with DNA are cell structures present in every prokaryotic cell. Most prokaryotic cells are surronded by cell wall made of peptidoglycan. Prokaryotic cells may have capsule, made of polysaccharids, the presence of which is often correlated with pathogenicity. Cell appendages like flagella and cilia ensure cell motility.
https://en.wikipedia.org/wiki/File:Prokaryote_cell_diagram.svg#file; LadyofHats, Public domain, via Wikimedia Commons

Eukaryotic cells are compartmentalized (Figure 23.)**.** Intracellular membranes divide the eukaryotic cell into compartments (organelles), each of which has a characteristic structure, biochemical composition, and function. The domain ***Eukarya*** consists of the major branches of ***protists, plants, fungi, and animals***. The most characteristic compartment of the eukaryotic cell is the ***nucleus*** (eu = "real"; karyon = "nucleus"). The main differences of the prokaryotic and eukaryotic cell organization are summarized in Table I.

	Prokaryotes	**Eukaryotes**
Cell size	~ 1-10 μm	~ 10-100 μm
Hereditary material	Stored in the nucleoid region; a single, circular chromosome	Stored in the nucleus, which is surrounded by double membrane; Linear chromosomes associated with histone proteins
Flow of genetic information	Transcription and translation in the cytoplasm; protein synthesis on 70S ribosomes	Transcription in the nucleus, translation in the cytoplasm; protein synthesis on 80S ribosomes
Structures in the cytoplasm	Few	Highly organized, inner membranes and cytoskeleton
Organization	Single cells	Single cells, colonies, multicellular organisms with specialized cells
Cell division	Fission	Mitosis (fission, budding) and meiosis

Table I. Comparison of Prokaryotes and Eukaryotes

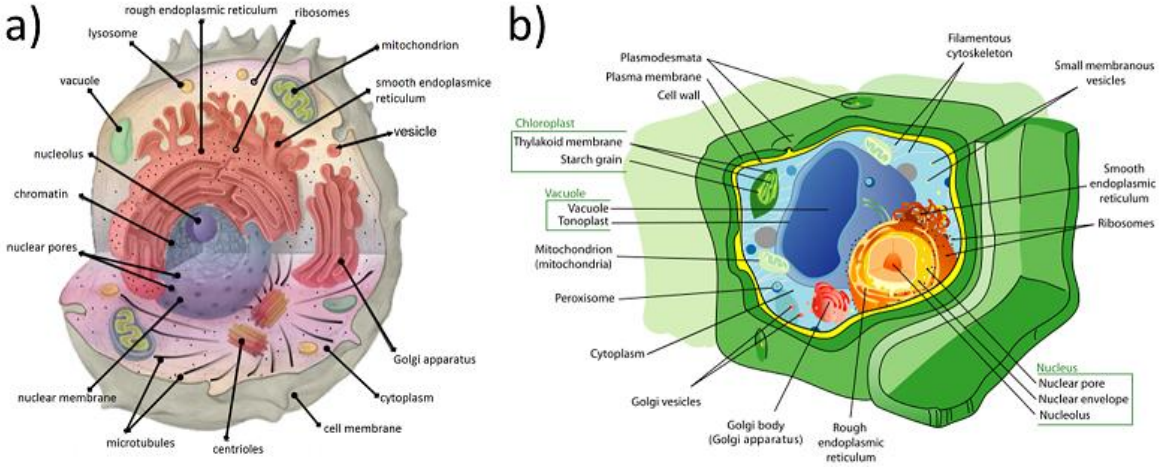


Figure 23. Eukaryotic cells

a) Organization of an animal cell. The plasma membrane and the cytosol belong to the most ancient cell structures, while the elements of cytoskeleton are typical of the eukaryotic cells. The nucleus and the mitochondria have two membranes. The endoplasmic reticulum, the Golgi apparatus, the peroxisome, and the elements of the endosomal-lysosomal compartment are made of or surrounded by a single membrane.
b) Plant cell. Note the cell wall, made of cellulose, the chloroplasts, and the large vacuole, which are typical structures of a plant cell.
https://commons.wikimedia.org/wiki/File:Cell-organelles-labeled.png
Bingbongboing, CC BY-SA 4.0 <https://creativecommons.org/licenses/by-sa/4.0>, via Wikimedia Commons
https://en.wikipedia.org/wiki/File:Plant_cell_structure-en.svg#file
Source: Self-made using Adobe Illustrator. (The original edited was also made by me, LadyofHats); Public domain

Overview of eukaryotic cellular organization and function

The ***nucleus*** encompasses the hereditary material and controls the cell's activity. Its structure and function will be discussed in detail later.

Ribosomes are molecular machines for protein synthesis. They consist of a large and a small subunit, both of which are made up of several RNAs and proteins. Ribosomes can bind to the ER or work free in the cytoplasm.

The ***endoplasmic reticulum*** (ER) is a continuous system of flattened membrane sacks and tubules. Proteins are made for insertion into cellular membranes or for secretion are synthesized on the ribosomes of the rough endoplasmic reticulum (***RER***). The smooth endoplasmic reticulum (***SER***) produces most of the lipid for the cell, is a storage place for Ca^{2+} ions, and contributes to detoxification.

The ***Golgi apparatus*** is a stack of flattened, membrane-bound sacks with many associated vesicles. It receives proteins from the endoplasmic reticulum, which are then modified here, and forwarded to their cellular destination or packed for secretion.

Lysosomes are small vesicles that contain degradative enzymes in acidic environment. They are sites for cellular digestion. Membrane vesicles, called ***endosomes***, deliver the substances destined for destruction to the lysosomes.

Mitochondria are double-membrane-built organelles, usually depicted as stiff, elongated cylinders. Mitochondria are "the powerhouses of the cell", because they produce most of the ATP for the cell. Mitochondria have their own DNA and ribosomes, which remind us to their evolutionary origin.

Peroxisomes are surrounded by a single membrane, and resemble lysosomes and endosomes morphologically. They contain enzymes that participate in oxidative reactions.

These enzymes might be at such a high concentration that they form a crystal in the peroxisome.

Eukaryotic cells have a ***cytoskeleton***. Actin filaments, microtubules, and intermediate filaments form a viscous and elastic cytoplasmic matrix to provide mechanical support for the cell. Actin filaments and microtubules are involved in various forms of cellular movement, as well.

Plant cells have some unique organelles. Plastids occur only in plant cells; the most characteristic for plants are the ***chloroplasts.*** Similarly to mitochondria, chloroplasts arose by endosymbiosis. Chloroplasts consist of complex membrane systems and carry out photosynthesis. ***Chromoplasts*** and ***leucoplasts*** store substances. ***Vacuoles*** are storage places for toxic by-products and waste material. Plant cells always have a ***cell wall*** whose main component is cellulose.

2.2 The nucleus

The ***evolutionary origin of the nucleus*** can be described as follows. In ancient prokaryotes, the hereditary material, a single DNA ring, was attached to the plasma membrane. The plasma membrane could have rearranged: formed an envelope around the DNA by invagination, and eventually pinched off completely. The transport between the nucleus and the cytoplasm was maintained through the nuclear pores. The perinuclear space became continuous with the lumen of the endoplasmic reticulum, and networks of intermediate filaments supported the nuclear envelope mechanically.

The nuclear envelope provides an advantage for the eukaryotes: it protects the chromatin from mechanical effects resulting from cytoskeletal rearrangements. Furthermore, the appearance of the nucleus has led to the ***spatial and temporal separation of the information transfer processes***, and the development of ***more complex regulation***.

In electron micrograph, the most prominent structures in an interphase (non-dividing) nucleus are the nuclear membrane with pores, heterochromatin, euchromatin, and the nucleolus (Figure 24).

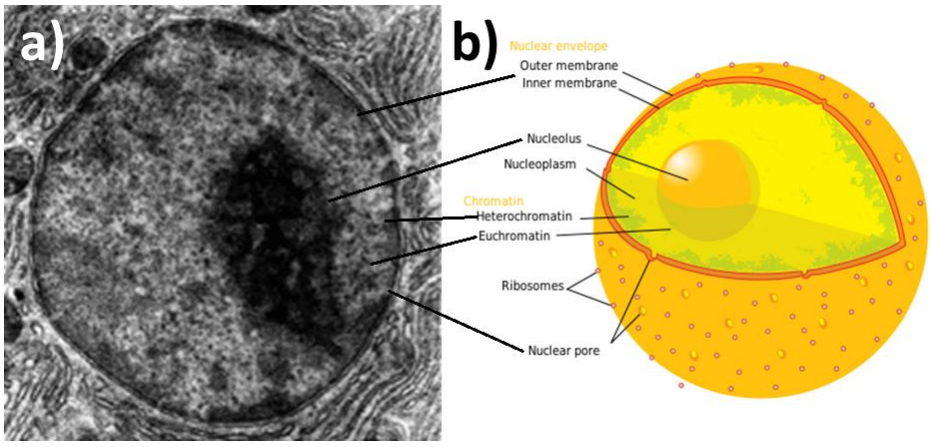


Figure 24. The cell nucleus in an EM image a) and schematic drawing b)

https://commons.wikimedia.org/wiki/File:Micrograph_of_a_cell_nucleus.png#file
https://commons.wikimedia.org/wiki/File:Diagram_human_cell_nucleus.svg
LadyofHats, Public domain, via Wikimedia Commons

The nuclear envelope consists of an ***outer membrane*** and an ***inner membrane*** that differ from each other both in composition and structure. There is a ***perinuclear space*** between the two membranes. Beneath the inner membrane, there is a meshwork of intermediate filaments called the ***nuclear lamina***. ***Nuclear pore complexes*** (Figure 25.) ensure selective, bidirectional traffic between the nucleoplasm and the cytoplasm. The largest space in the nucleus is occupied by ***chromatin***, whose structure will be discussed later. The ***nucleolus*** is the most conspicuous subnuclear structure seen in a eukaryotic nucleus. The rest of the nucleus is filled with so-called ***interchromatin domain*** that includes filaments, the nucleoplasm, and various macromolecular complexes.

The ***nuclear envelope is a selectively permeable barrier*** between the nuclear compartment and the cytoplasm. The outer nuclear membrane is not only continuous with the ER, but also shares its functions, e.g. it has ribosomes on its surface. The fibrous nuclear lamina supports the inner membrane and mediates its interaction with chromatin. ***Chromosomes are anchored*** to the inner membrane-nuclear lamina structure and occupy distinct territories within the nucleus.

Nuclear pores (Figure 25.) are huge protein complexes bridging both nuclear membranes. Besides being the sole ***route for communication*** between the nucleus and the cytoplasm, they act as ***diffusion barriers*** between the outer and inner membranes.

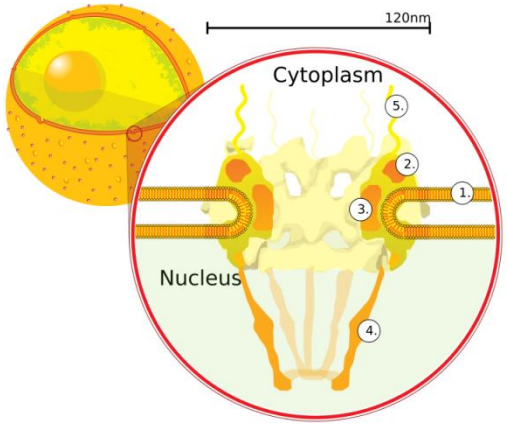


Figure 25. The structure of a nuclear pore

1. Nuclear envelope; 2. Outer ring; 3. Spokes; 4. Basket

The nucleolus has three functionally distinct regions:

1. Pars amorpha or ***fibrillar center***;
2. Pars fibrosa or ***dense fibrillar component***;
3. Pars granulosa or ***granular component***.

The nucleolus is formed around the ***nucleolus-organizer region (NOR)***. This is made up of chromosomal regions that carry the genes coding for rRNA. Fibrillar centers contain rRNA genes (rDNA) and the enzyme that transcribe them (RNA pol I). ***Precursors of ribosomal RNAs are made*** here. These ***precursors are then processed*** in the dense fibrillar components. Multitudes of rRNAs provide these parts of the nucleolus a fibrillar appearance. The granular component is rich in ribosomal proteins and pre-ribosomal particles. ***Ribosome subunits are assembled*** here.

2.3 Chromatin and chromosomes

DNA packaging in the chromatin

The genomes of eukaryotes are divided up into chromosomes. Each chromosome consists of a single, very long DNA molecule. ***In a human cell, 46 chromosomes consist of 2m DNA***, which must fit into a nucleus 6μm in diameter. To achieve this, DNA molecules must associate with proteins that fold and pack the DNA threads into more compact structures. ***The complex of DNA and proteins is called chromatin.*** Chromatin condensation changes during the cell cycle. In interphase cells chromatin forms loose loops, while it becomes highly condensed during cell division. ***Chromosomes*** are DNA molecules with their attendant proteins that ***move as independent units*** during mitosis and meiosis.

If chromatin is extracted from the cell and treated with differing concentration of salt, different degrees of compaction or condensation can be observed under the electron microscope (Figure 26.). The 30nm thick thread is the loosest natural form; the chromatin that resembles a bead necklace has been artificially decondensed. In the "bead-on-a-string" form of chromatin, the "beads" are the nucleosome core particles, and the "string" is the linker DNA between them. A ***nucleosome "bead"*** contains a protein core made of ***eight histone molecules***, and 146bp DNA. The DNA is wound around the histone octamer ~1.5-2x. Histone proteins protect this DNA from DNase enzymes. The size of the ***linker DNA is ~60bp***.

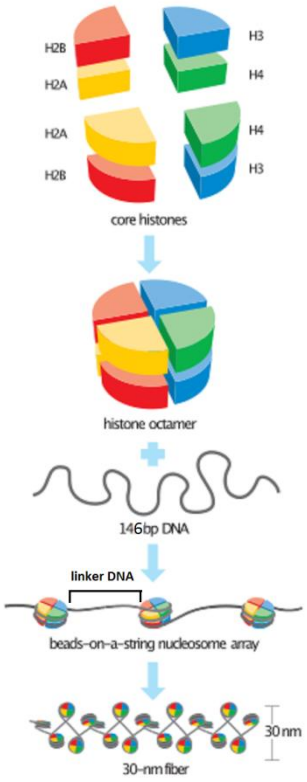


Figure 26. Basic units of chromatin structure

Chromatin can be experimantally decondensed using DNase enzyme and high salt concentration. For description of the components, see main text.

Based on: https://commons.wikimedia.org/wiki/File:Basic_units_of_chromatin_structure.svg;
David O Morgan, Attribution, via Wikimedia Commons

Histones are relatively small proteins that are rich in basic amino acids (lysine, arginine). They are typical of eukaryotic cells. Chromatin consists of histones and DNA in about a half to half ratio. ***There are five types of histones: H1, H2A, H2B, H3, and H4.*** In the nucleosome, two copies of each ***nucleosomal/core histones (H2A, H2B, H3, H4)*** are assembled into a histone octamer. Nucleosomal histones belong to the evolutionary most conserved proteins.

In interphase nuclei, most of the chromatin is in the form of a fiber with a diameter of about 30nm. The exact structure of the ***30nm chromatin fiber***, usually called ***solenoid,*** is still controversial. ***H1 or linker histone*** binds to linker DNA between nucleosomes and plays a role in organizing the 30nm fiber (Figure 27.). H1 is somewhat larger, and less conservative than the core histones.

From chromatin thread to metaphase chromosome

In addition to histones, chromatin consists of a variety of non-histone proteins, among others proteins involved in ***structural organization of chromosomes***. Interphase chromosomes are folded into a series of ***looped domains*** (Figure 27.), each containing 20-100kb nucleotide pairs of double helical DNA condensed into a 30nm fiber. Special regions of the DNA ***(scaffold attachment regions, SARs)*** associated by scaffold proteins anchor these loops to the ***chromosome scaffold***.

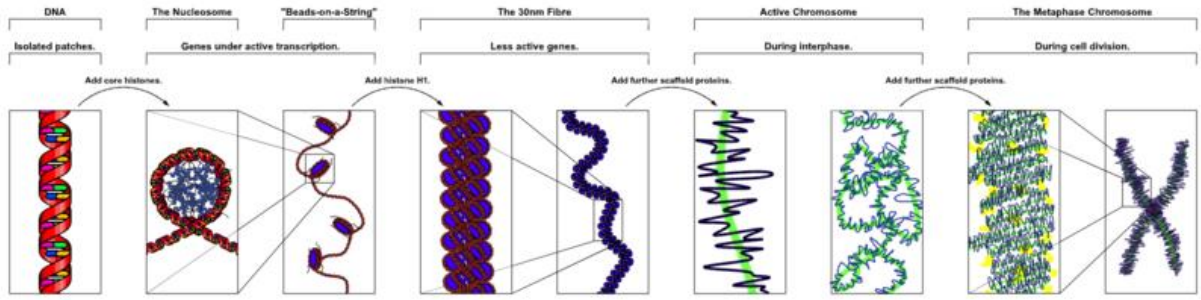


Figure 27. From DNA through chromatin to metaphase chromosome

Schematic representation of the gradual condensation of the hereditary material and the components involved at the different stages. For a more detailed explanation see the main text.
https://commons.wikimedia.org/wiki/File:Chromatin_Structures.png
Richard Wheeler at en.wikipedia, CC BY-SA 3.0 <http://creativecommons.org/licenses/by-sa/3.0/>, via Wikimedia Commons

Proteins identified in the chromosome scaffold are the enzyme topoisomerase *II* (topo *II*), condensins, and cohesins. ***Topoisomerase II*** makes and repairs double-stranded breaks, meanwhile sliding one DNA strand over the other. Topo *II* probably "untangles" the chromatin fibers that got tangled during replication. This enzyme is always present in the chromosome scaffold, but not as a structural building block. It is very dynamic *in vivo*, moving on and off of chromosomes quickly. The condensin and cohesin complexes are made of proteins responsible for the structural maintenance of chromosomes, and other proteins. ***Condensins*** are needed to regular chromosome condensation. ***Cohesin*** complexes are loaded onto the chromosomes during DNA replication, and hold sister chromatids together until the anaphase of cell division.

Relationship between chromatin structure and function

In a typical interphase nucleus, about 10% of the genome is always in a highly condensed form, called constitutive heterochromatin. ***Constitutive heterochromatin*** is composed of DNA sequences that are ***never transcribed*** in any cell type. Several structural

proteins are involved in forming this particularly compact chromatin structure. Satellite DNA at chromosomal ***centromeres*** and ***telomeres*** are packed into constitutive heterochromatin. Telomeres are anchored to the nuclear lamina and form the so-called ***perinuclear heterochromatin***.

In a typical eukaryotic nucleus, about 10% of the chromatin is ***euchromatin***. Euchromatin contains genes that are being ***actively transcribed***. It has loose structure.

Facultative heterochromatin is a more or less condensed region of the genome. It ***contains genes, but it is not actively transcribed*** in every cell. The ***X chromosome*** is the classical example of facultative heterochromatin in mammals (Figure 28.). These forms of chromatin are different structurally, functionally and biochemically. The quality and quantity of proteins associated with DNA is different in the different chromatin types.

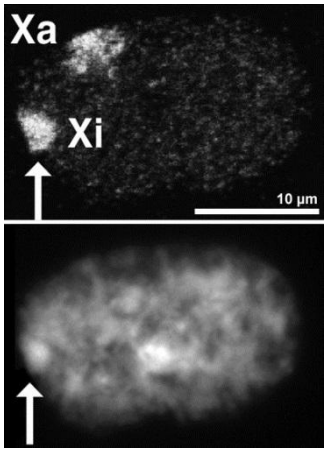


Figure 28. Nucleus of a female amniotic fluid cell

Top: Both X chromosomes are detected by FISH. Xa is the active, Xi is the inactive copy of chromosome X. The inactive X chromosome is also called ***Barr body***. Bottom: The same nucleus stained with DAPI. The arrow points to the Barr body.
https://commons.wikimedia.org/wiki/File:Sd4hi-unten-crop.jpg; User:Dietzel65, Steffen Dietzel, CC BY-SA 3.0 <https://creativecommons.org/licenses/by-sa/3.0CC BY-SA 3.0 Creative Commons Attribution-Share Alike 3.0>, via Wikimedia Commons

Chromosomal landmarks

Features such as ***size, arm ratio, heterochromatin, number and position of thickenings, number and location of the nucleolar organizers, and banding pattern identify the individual chromosomes*** within the set that characterize a species.

The number of chromosomes is characteristic of a species (Table II.). It is the product of the haploid number and the number of sets. The haploid number (n) is the number of chromosomes in the basic genomic set. In higher organisms (animals and plants), somatic cells have two sets of chromosomes; such cells are called diploid (2n). The range of haploid chromosome number varies widely, from two to several hundreds.

Common name	Scientific name	Number of chromosome pairs	Common name	Scientific name	Number of chromosome pairs
Garden onion	*Allium cepa*	8	Cattle	*Bos taurus*	30
Rice	*Oryza sativa*	12	Horse	*Equus caballus*	32
Wheat	*Triticum aestivum*	21	Chicken	*Gallus domesticus*	39
Potato	*Solanum tuberosum*	24	Carp	*Cyprinus carpio*	52

Table II. Chromosome number in some economically important species

The chromosomes of a single genome may differ considerably in size. For an example, see the human chromosome set later, in Figure 30.

Chromosome morphology can be described best on metaphase chromosomes. Chromosomes in the metaphase stage of cell division represent the most condensed form of chromatin. Each mitotic chromosome consists of two ***sister chromatids*** that are held together at the primary constriction called the ***centromere***. Both chromatids contain a single DNA strand due to a prior DNA duplication in S phase. Chromosomes are characterized by size and the ***position of the centromere*** relative to the arms. When the centromere is located asymmetrically, the chromatids can be divided into ***short (P) and long (Q) arms***. According to the position of the centromere, chromosomes can be classified as metacentric, submetacentric, acrocentric, telocentric (Figure 29.)

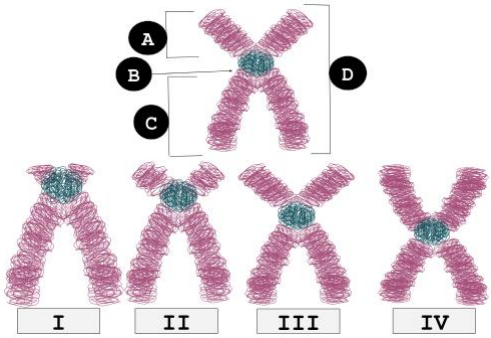


Figure 29. Classification of chromosomes

A: Short arm (P); B: Centromere; C: Long arm (Q); D: Sister chromatid.
I. Telocentric; II. Acrocentric; III. Submetacentric; IV. Metacentric.
https://commons.wikimedia.org/wiki/File:Centromere_Placement.svg
Fockey003, CC BY-SA 4.0 <https://creativecommons.org/licenses/by-sa/4.0>, via Wikimedia Commons

A traditional way to distinguish mitotic chromosomes is to stain them with dyes that produce a striking and reliable pattern of bands. The position and sizes of DNA ***bands are constant and specific*** to the individual chromosomes.

One of the basic methods is the application of Giemsa reagent, a DNA-specific dye, after mild proteolytic digestion of the metaphase chromosomes. Giemsa produces patterns of ***light-staining (G-light)*** regions and ***dark-staining (G-dark)*** regions. The G-dark regions are packed more densely than the G-light regions, thus there is a higher density of DNA in them to take up the stain. The dark regions tend to be heterochromatic, late-replicating and AT rich. In contrast, G-light regions contain most of the active genes; they tend to be early replicating, and relatively GC-rich.

Giemsa staining is used to describe the human ***karyotype*** (Figure 30.). The karyotype is the set of chromosomes, typical of a species. The human karyotype contains ***22 pairs of autosomal chromosomes and one pair of sex chromosomes***. The common karyotype for females is denoted ***46, XX***; for males ***46, XY***. Comparison of a ***karyogram***, the chromosome set of an individual, with the karyotype can reveal chromosome aberrations.

Figure 30. Karyotype of a human male

Metaphase chromosomes, stained with Giemsa and arranged according to size.
The last pair is the sex chromosomes.
https://commons.wikimedia.org/wiki/File:NHGRI_human_male_karyotype.png
Courtesy: National Human Genome Research Institute, Public domain, via Wikimedia Commons

Centromeres and telomeres

The centromere or primary constriction is ***the region where the chromosome becomes attached to the spindle***; therefore, it is at the heart of all chromosomal movements in mitosis and meiosis. It has an important role in monitoring the attachment of the chromosomes to the kinetochore microtubules, as well. Centromeres are ***nucleoprotein structures***, in which both DNA and proteins play an important role. Centromeric DNA usually consists of groups of repetitive sequences; in case of vertebrates, it is the highly repetitive ***satellite DNA***. The chromatin of centromeres is entirely composed of ***constitutive heterochromatin***, to which ***kinetochore proteins*** are associated.

Centromere types: ***point centromere*** (binds to one kinetochore microtubule; e.g.: *S. cerevisiae*), ***regional centromere*** (2-20 microtubules; e.g.: *S. pombe*, *Drosophila*, human), ***holocentromere*** (centromere activity along the whole chromosome; e.g.: *C. elegans*).

Telomeres are the ends of chromosomes. The ends of chromosomes represent a challenge to the replication mechanism; to solve this, cells package the chromosome ends into ***specialized structures***. This is necessary to distinguish the ends of chromosomes from simple DNA breaks, and to replicate the chromosomal DNA out to the very end.

Telomeres are composed of ***many repeats of short DNA sequences (5'TTAGGG3').*** Several proteins are involved in forming the telomere structure. Some proteins bind directly to the repeated sequence, others to other proteins.

Special chromosomes

Polytene chromosomes

Certain insect cells replicate their DNA several times without a following separation and cell division. This type of division is called ***endomitosis***. In the ***salivary gland cells of Drosophila larvae***, 1024 (2^{10}) identical strands of chromatin are lined up side by side, forming a ***giant chromosome***. These ***polytene chromosomes*** in the insect salivary glands are suitable for studying the structure of interphase chromosomes (Figure 31.).

Polytene chromosomes have distinct, alternating ***dark bands and light interbands***. The chromatin is much more condensed in the dark bands than in the interbands. Those sites of the giant chromosomes where intense transcription takes place swell and form so-called ***"puffs"*** (swollen parts). Changes in gene expression can be monitored by changing of the pattern of puffs. Polytene chromosomes are the ***most typical in the secretory tissue of***

dipteran insects, but can also be found in other taxa as well. In plants, they are usually found in the endosperm.

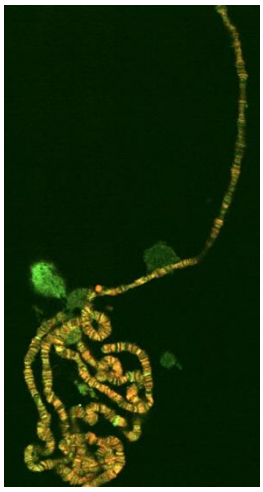


Figure 31. Polytene chromosome from fly salivary gland

https://commons.wikimedia.org/wiki/File:Polyten_chromosome.jpg
LPLT, Public domain, via Wikimedia Commons

Sex chromosomes

Most animals and many plants show sexual dimorphism. In most of these species, sex is determined by a special pair of ***sex chromosomes***. Both in humans and *Drosophila*, females have a pair of identical sex chromosomes called the ***X chromosome***. Males carry a nonidentical pair, consisting of one X and one ***Y chromosome***. The Y chromosome is much smaller than the X. The female is said to be the ***homogametic sex***, because in the eggs there is only one type of sex chromosomes. Males produce sperms with X and Y chromosome in a half to half ratio, therefore, the male is called the ***heterogametic sex***.

The mechanism of sex determination in *Drosophila* differs from that of in human (or mammals, in general). In *Drosophila*, the ***number of X chromosomes*** determines sex, while in mammals the ***presence of the Y chromosome***.

The nonidentical sex chromosomes bear homologues regions called ***pseudoautosomal regions*** (Figure 32.). These pseudoautosomal regions pair up in meiosis.

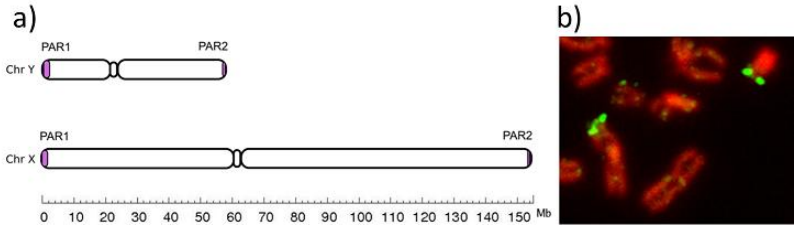


Figure 32. Pseudoautosomal regions of sex chromosomes

a) Schematic representation of the human Y and X chromosomes. They differ largely in size, and also in DNA content. The pseudoautosomal regions PAR1 and PAR2 are used for meiotic pairing.
b) Detail of a human metaphase spread. FISH probe (green) shows the PAR1 regions of chromosome X and Y.
https://commons.wikimedia.org/wiki/File:Pseudoautosomal_region.png; Ashwin Kelkar , Vivek Thakur , Ramakrishna Ramaswamy , Deepti Deobagkar, CC BY 2.5 <https://creativecommons.org/licenses/by/2.5>, via Wikimedia Commons
https://commons.wikimedia.org/wiki/File:A_region_in_the_pseudoautosomal_region_of_the_short_arms_of_the_X-_and_Y-chromosome.jpg; Steffen Dietzel, Dietzel65, CC BY 3.0 <https://creativecommons.org/licenses/by/3.0>, via Wikimedia Commons

Vascular plants show a variety of sexual arrangements. ***Dioecious*** species (Figure 33.) show animal-like sexual dimorphism, with female plants bearing flowers with ovaries only, and male plants bearing flowers with anthers only. Some dioecious plants have a nonidentical pair of chromosomes for sex determination. For example, the plant *Silene latifolia* has the XY system.

Figure 33. Two dioecious plant species

a) *Aruncus dioicus* (goat's beard, buck's beard, bride's feathers), male inflorescence.
b) *Silene latifolia* (white campion; formerly called *Melandrium album*)
https://commons.wikimedia.org/wiki/File:Aruncus_dioicus_1549.JPG;
Walter Siegmund, CC BY-SA 3.0 <https://creativecommons.org/licenses/by-sa/3.0>, via Wikimedia Commons
https://commons.wikimedia.org/wiki/File:Silene_latifolia_9638.JPG
Walter Siegmund, CC BY-SA 3.0 <https://creativecommons.org/licenses/by-sa/3.0>, via Wikimedia Commons

B chromosomes

Although the chromosome number of species has stabilized to a large extent over the course of evolution, ***B chromosomes*** are an exception to this general rule. They are also called ***supernumerary, accessory, dispensable or lineage-specific chromosomes***, and occur in many animal, plant and fungal species (Figure 34.).

B chromosomes are ***not essential***, and are lacking in some (usually most) individuals of the species. Most B chromosomes are mainly or entirely ***heterochromatic***, and smaller than A chromosomes. Presumably, they derived from heterochromatic segments of normal chromosomes in the remote past.

Among plants, they are most common in grasses (in the family *Poaceae* also called *Gramineae*). It was observed that an increase in the frequency of B chromosomes is usually associated with decreased fertility and viability. In other cases, B chromosomes provide some adaptive advantage.

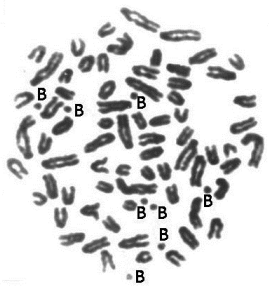


Figure 34. Metaphase spread of the Siberian Roe deer (*Capreolus pygarus*) with B chromosomes

https://commons.wikimedia.org/wiki/File:Metaphase_spread_of_the_Siberian_Roe_deer_(Capreolus_pygargus).jpg
Graphodatsky et al., CC BY 2.0 <https://creativecommons.org/licenses/by/2.0>, via Wikimedia Commons

2.4 Types of cell division

Stages of the cell cycle

The cell cycle (Figure 35.) comprises a series of events that take place in a cell from one cell division to the next one. For early scientists, cells appeared to be active only during division (***M phase***), so the rest of the cell cycle was considered as a resting phase (***interphase***). However, cells carry out different activities in interphase, as well. The interphase can be further subdivided into ***G1, S,*** and ***G2*** phases.

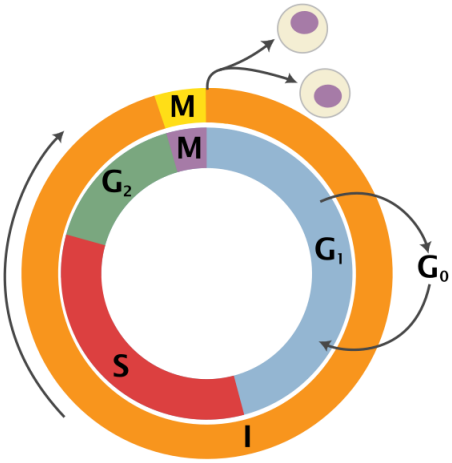


Figure 35. The cell cycle

The outer circle presents a simple division of the cell cycle into M phase and interphase. The inner circle shows the stages of cell cycle in regularly dividing cells. The drawing depicts the stages according to the length of their real period of time. Differentiated cells do not divide; they exit the cell cycle and stay in the G0 resting phase. The other cell cycle stages are described in the main text.
https://commons.wikimedia.org/wiki/File:Cell_Cycle_2-2.svg
Cell_Cycle_2.svg: *Cell_Cycle_2.png:Zephyris at en.wikipediaderivative work: Beaoderivative work: Histidine, CC BY-SA 3.0 <http://creativecommons.org/licenses/by-sa/3.0/>, via Wikimedia Commons

During the ***G1 phase*** (first ***gap*** phase) cells grow, monitor their inner and outer environment, and prepare for DNA synthesis. In G1, each chromosome contains a single DNA strand. The chromosomal DNA is replicated during ***S (synthesis) phase***. During the ***G2 phase*** (second gap phase), cells prepare for division. As a result of the previous DNA replication, now chromosomes consist of two side-by-side units called sister chromatids that are held fast together. In ***M phase***, first sister chromatids are separated and segregated into two daughter nuclei (***mitosis***), and then the cell divides (***cytokinesis***).

Mitosis

Mitosis is the nuclear division associated with the asexual division of cells. In multicellular organisms, ***somatic cells*** (the cells of the body) divide by mitosis. Mitosis gives rise to ***genetically identical cells*** in which the total ***number of chromosomes is maintained***. Therefore, mitosis is also known as ***equational division***. Mitosis is preceded by a single round of DNA replication, and is ***resulted in two daughter nuclei***.

Mitotic events are subdivided into phases, such as ***prophase, metaphase, anaphase and telophase*** (Figure 36.).

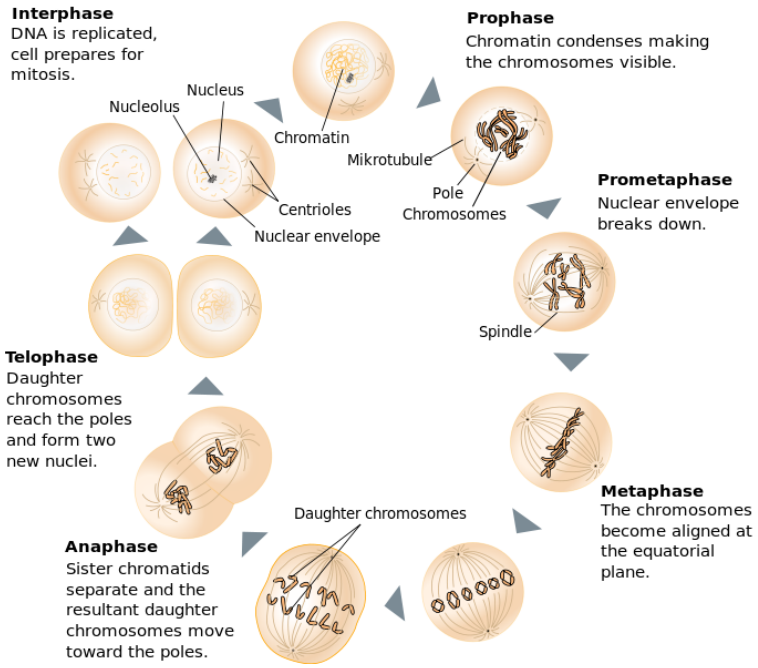


Figure 36. Schematic diagram of mitosis

https://commons.wikimedia.org/wiki/File:Mitosis_schematic_diagram-en.svg
Original: Jpablo cad and juliana osoriotranslation: MattDerivative work: M3.dahl, CC BY-SA 3.0 <https://creativecommons.org/licenses/by-sa/3.0>, via Wikimedia Commons

Prophase: ***Chromosomes condense*** (condensins), and pairs of sister chromatids become visible. They are held together by cohesin complexes. ***Centrosomes*** (animal) that have replicated during S phase ***move apart and form two new poles***. Between them a ***mitotic spindle*** begins to form that consists of two radial arrays of dynamic microtubules. Finally, the ***nuclear envelope breaks down***. Sometimes this stage is distinguished as ***prometaphase***.

Metaphase: The kinetochore microtubules of the spindle attach to chromosomes. ***Chromosomes line up at the equatorial plate of the cell***. Chromosomes achieve ***bipolar orientation***: sister kinetochores are attached to opposite poles. Proteins sitting on the kinetochore have a role in controlling the metaphase-anaphase transition.

Anaphase: Cleavage of the cohesin complex allows cells to separate sister chromatids that become daughter chromosomes. In ***anaphase A***, ***sister chromosomes are moved to the poles*** due to depolymerization of the kinetochore microtubules. In ***anaphase B***, the two ***spindle poles move*** further apart.

Telophase: As chromosomes have arrived to the poles, they ***decondense***, the spindle disassembles, and a ***nuclear membrane reforms*** around each nucleus.

Finally, organelles are distributed in ***cytokinesis***, and two daughter cells are formed.

Meiosis

Meiosis is an ancient Greek word, meaning “lessening”. In meiosis, ***a single round of DNA replication is followed by two divisions*** (Figure 37.). During the first, ***reductional division***, homologous chromosomes of maternal and paternal origin segregate. The second, ***equational division*** is analogous to mitosis: it separates sister chromatids. After two divisions, ***meiosis produces four haploid daughter cells from a diploid meiocyte***.

In humans and other ***animals,*** meiosis takes place in the gonads, and the products of meiosis are the ***gametes: sperms and eggs***. In flowering ***plants***, meiosis takes place in the anthers and ovaries, and the products of meiosis are ***meiospores***, which eventually give rise to gametes. In sexually reproducing organisms, this special type of cell division ***maintains the ploidy*** from generation to generation. Further, meiosis ***generates genetic diversity*** via recombination.

Of the meiotic stages, the prophase of the first division is the longest and most complex. This stage will be discussed in detail, because several meiosis specific events take place during prophase I.

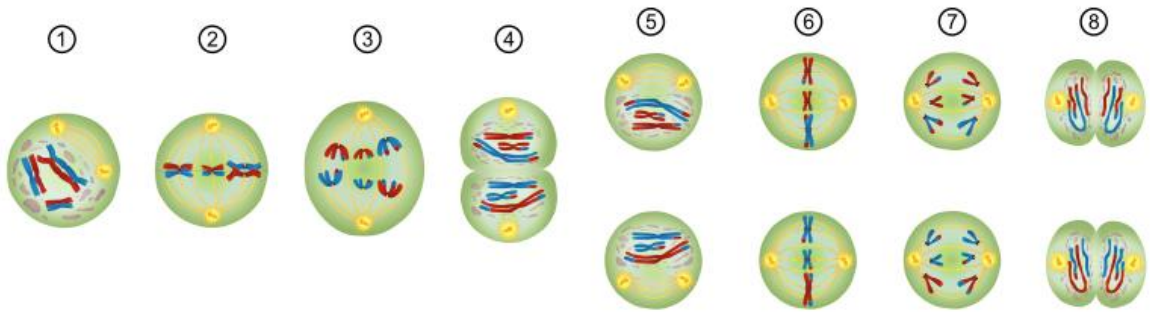


Figure 37. Schematic diagram of the stages of meiosis

1: Prophase I; 2: Metaphase I; 3: Anaphase I; 4: Telophase I and Cytokinesis;
5: Prophase II; 6: Metaphase II; 7: Anaphase II; 8: Telophase II and Cytokinesis
https://commons.wikimedia.org/wiki/File:Meiosis_Stages_-_Numerical_Version.svg
Ali Zifan, CC BY-SA 4.0 <https://creativecommons.org/licenses/by-sa/4.0>, via Wikimedia Commons

PROPHASE I

Leptotene – "thin threads": Chromosomes begin to ***condense and pair***. By the end of this stage, homologous chromosomes are aligned in a distance of ~400nm. A meiosis specific event, the ***meiotic recombination*** starts with the formation of double-stranded DNA breaks (Figure 38.).

Zygotene – "paired threads": Homologous chromosomes come together in a process known as ***synapsis***. This is linked with the assembly of the ***synaptonemal complex*** (Figure 38.). In many species, telomeres are clustered together at a spot on the nuclear envelope, giving rise to a ***"bouquet"*** arrangement of chromosomes.

Pachytene – "thick threads": Synapsis is complete; homologous chromosomes are joined together by a fully developed ***synaptonemal complex (SC;*** Figure 38.). Crossover sites "mature" into ***chiasmata.***

In meiosis, the homologous chromosomes are held together the most tightly during pachytene. The SC is a tripartite structure, formed exclusively in meiotic cells. It consists of two ***lateral elemens, transverse filaments*** connecting the lateral elements, and a ***central element***. The SC is accompanied by ***early and late nodules***, the latter of which are thought to contain proteins involved in meiotic recombination. The transition of crossover structures into chiasmata, the cytologically visible sites of meiotic recombination, occurs within the SC.

Diplotene – "two threads": The ***synaptonemal complex disintegrates***. Sister chromatids remain closely associated, whereas ***homologous chromosomes "repel"*** each other, held only by chiasmata.This stage can last for years in the females of some species. Where the level of transcription is particularly elevated, chromosomes adopt a special structure known as ***lampbrush chromosome***.

Diakinesis – "moving through": Diakinesis corresponds to the prometaphase of meiosis I. The nuclear envelope breaks down. Chromosomesare in their most condensed state.

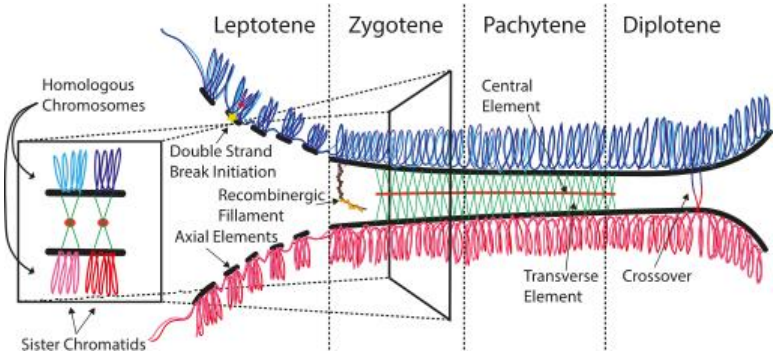


Figure 38. Schematic representation of the synaptonemal complex at different stages during Prophase I

https://commons.wikimedia.org/wiki/File:Synaptonemal_Complex.svg
Daniel Wells, CC BY-SA 4.0 <https://creativecommons.org/licenses/by-sa/4.0>, via Wikimedia Commons

METAPHASE I

The bivalents are aligned at the metaphase plate. Each of the homologues is oriented to one of the poles of the meiotic spindle.

ANAPHASE I

Homologues separate and move to opposite poles.

TELOPHASE I and INTERPHASE

These stages do not occur in every species. Where observed, they are characterized by chromosome decondensation and reformation of the nuclear envelope. ***There is no DNA replication between the meiotic divisions.***

MEIOSIS II

It consists of pro-, meta-, ana-, and telophase, which are essentially similar to the corresponding phases of mitosis.

During the meiotic divisions, the proper kinetochore orientation and gradual breakdown of sister chromatid cohesion are essential factors to the correct segregation of chromosomes.
For a summary and comparison of the two types of cell division, see Table III.

Mitosis	Meiosis
In somatic cells	In cells in the sexual cycle
Single DNA replication followed by one division that results in two daughter cells	Single DNA replication followed by two divisions that result in four daughter cells
Chromosome number is maintained	Chromosome number is halved
Cell undergoing mitosis can be diploid or haploid	Cell undergoing meiosis is diploid
No pairing of homologs	Full synapsis of homologs at prophase I
No crossover	At least one crossover per homologous pair
Daughter cells are genetically identical	Promotes genetic variation of the daughter cells
Sister chromatids separate at anaphase	Homologs separate at anaphase I, sister chromatids separate at anaphase II

Table III. Comparison of mitosis and meiosis

3. The Structure and Function of Nucleic Acids

3.1 The structure of DNA

DNA is made up of four different building blocks, the four ***nucleotides: dAMP, dGMP, dCMP, dTMP*** (Figure 39.). Each nucleotide contains three components: ***phosphate***, a ***sugar*** called ***deoxyribose***, and one of the four ***nitrogenous bases – adenine, guanine, cytosine or thymine***. ***Purine*** nucleotides contain bases that have a double-ring structure, whereas bases in ***pyrimidine*** nucleotides have a single-ring structure.

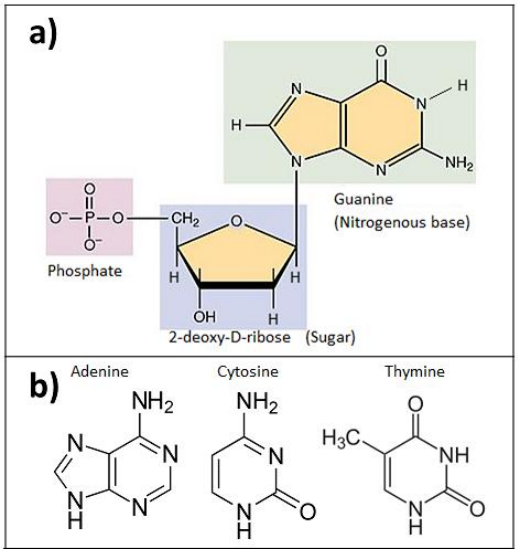


Figure 39. Building blocks of DNA

a) dGMP (deoxyguanosine-monophosphate), one of the four nucleotides that build up the DNA. The different types of compounds that form a nucleotide are highlighted.
b) Instead of guanine, adenine (purine-base), cytosine or thymine (pyrimidine bases), can stand in the DNA.

Biochemical studies and X-ray diffraction experiments have largely contributed to the elucidation of the structure of DNA. Studying a large selection of DNAs, Erwin Chargaff established empirical rules about the base composition of DNA: 1. A+G=T+C (The total amount of purine bases equals that of the pyrimidine bases.); 2. T=A and C=G (The molar ratio of thymine is the same as that of the adenine; same for cytosine and guanine.).

In X-ray diffraction experiments, X rays are fired at the DNA fibers, and the scatter of X rays is observed by catching them on photographic film. The angle of scatter represented by each spot on the film gives information about the position of an atom or certain groups in the DNA molecule. The X-ray data suggested that DNA is a long and skinny, probably spiral-like molecule.

The 3D structure of DNA is composed of two side-by-side chains, also called ***strands***, of nucleotides twisted into a ***double helix*** (Figure 40.). The structure resembles a spiral staircase. The ***backbone of each strand*** is formed of alternating ***phosphate and sugar*** units connected by ***phosphodiester linkages***. A phosphodiester bond connects the 5’ carbon atom of one deoxyribose to the 3’ carbon atom of the adjacent deoxyribose. Thus, each strand has a ***5’-to-3’ direction or polarity***. In the double-stranded DNA molecule, the two strands are in opposite or ***antiparallel orientation***.

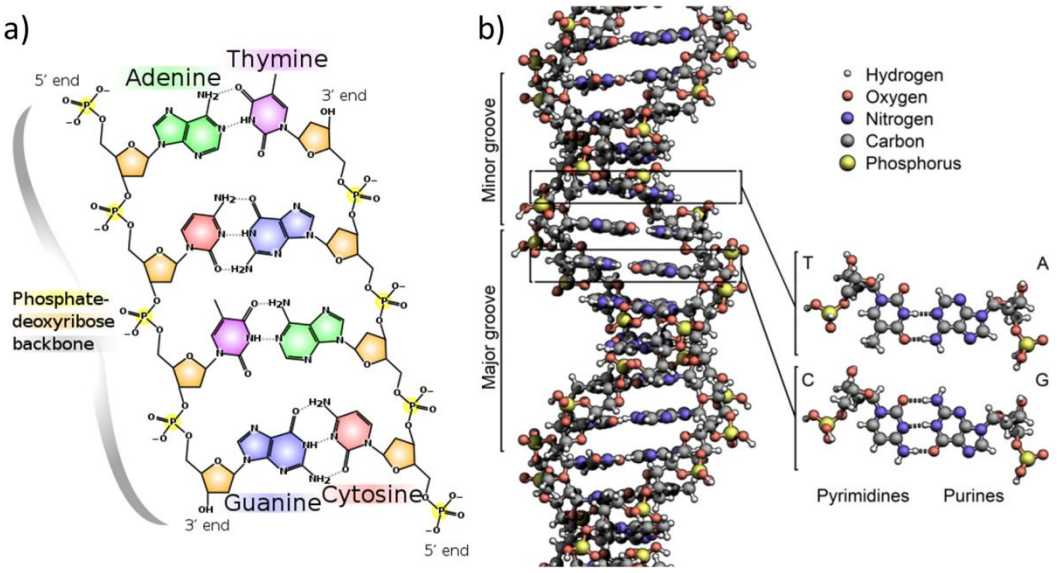


Figure 40. DNA models

a) Representation empasizing the nature of chemical bonds between the components
b) Spatial representation of the double helix
https://commons.wikimedia.org/wiki/File:DNA_chemical_structure.svg
Madprime (talk · contribs), CC0, via Wikimedia Commons
https://commons.wikimedia.org/wiki/File:DNA_Structure%2BKey%2BLabelled.pn_NoBB.png
Zephyris, CC BY-SA 3.0 <https://creativecommons.org/licenses/by-sa/3.0>, via Wikimedia Commons

The double helix has two grooves running in a spiral: the ***major groove*** and the ***minor groove***. DNA is a right-handed helix. The ***"steps"*** of the spiral staircase are ***base pairs***, which are flat planar structures at the center of DNA. Each base pair is built of one purine and one pyrimidine base, paired according to the following rule: ***G pairs with C, and A pairs with T***. The bases are held together by ***hydrogen bonds***: the G-C pair has three, whereas the A-T pair has two of them. ***DNA is a double helix composed of two nucleotide chains held together by complementary paring of A with T and G with C.***

3.2 Replication

Overview of DNA replication

DNA synthesis is catalyzed by ***DNA polymerase*** enzymes. The ***chain-elongation*** reaction takes place ***in 5' to 3' direction***; DNA polymerases add deoxyribonucleotides to the 3' end of a growing nucleotide chain (Figure 41.). A ***single stranded DNA*** that has been exposed by localized unwinding of the double helix ***serves as a template***. The substrates for DNA polymerases are ***dATP, dGTP, dCTP, and dTTP***. In *E. coli*, two DNA polymerase enzymes, ***pol I and pol III***, are involved in replication. DNA polymerases can extend a chain but cannot start a chain. DNA synthesis must be initiated by a ***primer***, which is a short chain of nucleotides that binds to the template strand to form a short duplex DNA.

DNA polymerases are capable of ***proofreading*** their job, that is, they ***remove misincorporated nucleotides*** from the synthesized strand. ***This is the key to the faithful copying of the genetic material.***

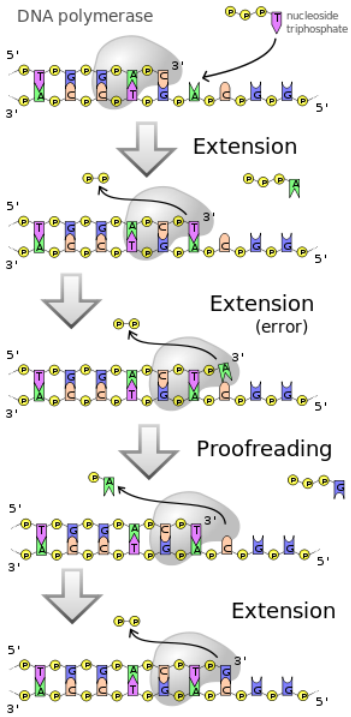


Figure 41. Diagram of DNA polymerase extending a DNA strand and proofreading

https://commons.wikimedia.org/wiki/File:DNA_polymerase.svg
I, Madprime, CC0, via Wikimedia Commons

The copying mechanism of DNA is called ***semiconservative***. In semiconservative replication, the double helix of each daughter DNA molecule contains one strand from the original DNA molecule and one newly synthesized strand.

The ***replication fork*** is the site at which the double helix is unwound to produce two single strands that serve as templates for copying DNA (Figure 42.).

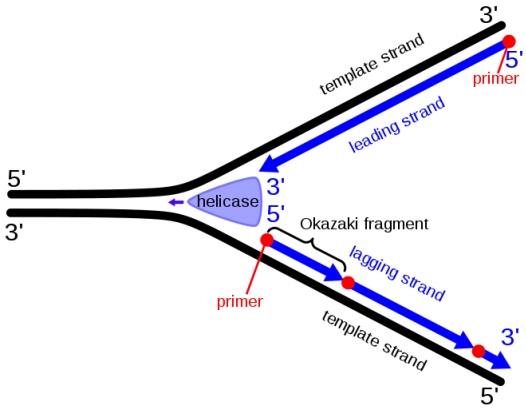


Figure 42. Replication fork

The helicase unwinds the double helix for the polymerase. The leading strand is continuously extended from a single primer, while the lagging strand is synthesized in Okazaki fragment "units", each of which is started with a primer.
https://commons.wikimedia.org/wiki/File:Okazaki_fragment_EN.svg
Andi schmitt, CC BY-SA 4.0 <https://creativecommons.org/licenses/by-sa/4.0>, via Wikimedia Commons

DNA pol *III* works at the replication fork. However, because DNA polymerases always add nucleotides at the 3' growing tip, only one strand can serve as a template in the direction of the replication fork. For this strand, ***synthesis can take place continuously***; the new strand synthesized on this template is called the ***leading strand***. Synthesis on the other template is in the "wrong" direction, because it has to move away from the replication fork. This other strand, called the ***lagging strand***, must be ***synthesized in short segments***.

Synthesis of both the leading strand and each DNA segment of the lagging strand must be initiated by a primer. Primers are synthesized by the ***primase*** enzyme (see on Figure 43.), which is a type of RNA polymerase. Primase synthesizes short (~8-12 nucleotides) ***RNA primers***. The primers are then extended as a DNA chain by ***DNA polymerase III***. The short (1000-2000 nucleotides) stretches of newly synthesized DNA of the lagging strand are called ***Okazaki fragments***. Later ***DNA polymerase I*** removes the RNA primers and fills in the resulting gaps with DNA. Finally, ***DNA ligase*** joins Okazaki fragments by catalyzing the formation of a phosphodiester bond between the 5'-phosphate end of one fragment, and the adjacent 3'-OH group of another fragment. It is the only enzyme that can seal DNA chains.

The replisome

The two hallmarks of DNA replication are ***accuracy and speed***. In *E. coli*, a replication fork moves at a rate of 1000 nucleotides per second. To accomplish this, DNA polymerase works as a part of a large ***nucleoprotein complex*** that coordinates the activities at the replication forks. This complex, called the ***replisome***, is an example of a ***"molecular machine"*** (Figure 43.).

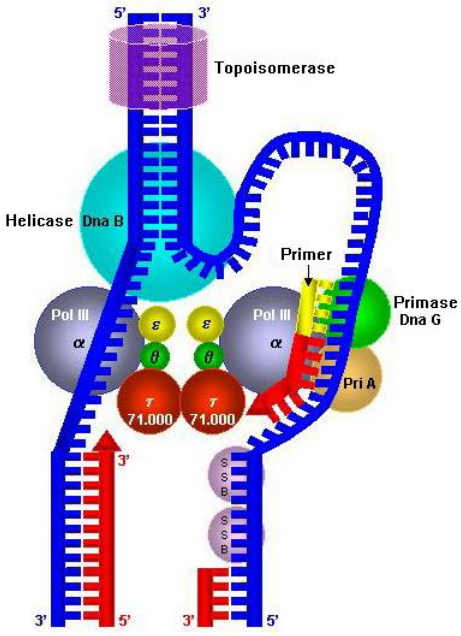


Figure 43. Model of the *E. coli* replisome

Both pol *III* and the primosome contain more subunits than indicated. (Pol III α: polymerase activity, ε: proofreading, θ: supports ε, τ: joins pol *III* molecules. PriA: primosome assembly. SSB: single-strand binding proteins.)

https://commons.wikimedia.org/wiki/File:E._coli_replisome.png

César Benito Jiménez with English-language re-captioning and other modifications by Stigmatella aurantiaca, CC BY-SA 2.5 <https://creativecommons.org/licenses/by-sa/2.5CC BY-SA 2.5 Creative Commons Attribution-Share Alike 2.5>, via Wikimedia Commons

The catalytic core of DNA pol *III* is part of a large complex, called the ***pol III holoenzyme***. Pol *III* ***works as a dimer***, one catalytic core handles the synthesis of the leading strand, while

the other handles lagging strand synthesis. There are many accessory proteins in the pol *III* holoenzyme; some of them coordinate the synthesis of the leading and lagging strand, others keep pol *III* attached to the DNA molecule (***sliding clamp***). Extra-twisted regions accumulate ahead of the replication fork as the parental strands separate. ***Topoisomerase*** enzymes, like e.g. gyrase, remove these regions, allowing the DNA to relax. ***Helicases*** are enzymes that disrupt the hydrogen bonds that hold the two strands of the double helix together. The unwound DNA is stabilized by ***single-strand binding (SSB) proteins***, which bind to single stranded DNA and prevent the duplex from re-forming.

Differences of the prokaryotic and eukaryotic replication

DNA replication in eukaryotes is somewhat more complex than in prokaryotes. One reason for the higher complexity is the higher ***complexity of the eukaryotic template***. Eukaryotic chromosomes exist in the nucleus as ***chromatin.*** Eukaryotic replisomes do not only have to copy the parental strands, but also disassemble the nucleosomes in the parental strand and reassemble them in the daughter molecules.

Assembly of the replisome begins at precise sites on the chromosomes, called origins. Replication proceeds in both directions from a single origin on the circular prokaryotic chromosome. In order to replicate the much larger eukaryotic genomes quickly, ***eukaryotic chromosomes have many replication origins*** (Figure 44.). Furthermore, eukaryotes have to integrate DNA replication into the cell cycle: ***all the chromosomes replicate once, at the same time, during the S phase***.

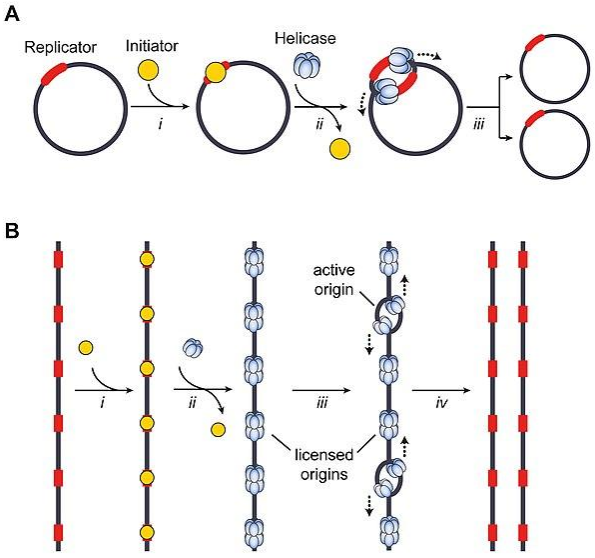


Figure 44. Origins of DNA replication

a) The smaller prokaryotic genome replicates from a single origin.
b) Rapid replication of the larger eukaryotic genome requires several origins of replication. After origin recognition and activation, the DNA opens at several sites along the chromosome.
https://commons.wikimedia.org/wiki/File:Origins_of_DNA_replication_Figure_1.jpg#file
https://creativecommons.org/licenses/by/4.0/deed.hu

Prokaryotic chromosome is replicated fully, starting from a single replication origin. However, replication of the two ends of linear DNA molecules, that is the telomeres of chromosomes, poses a problem. Once the primer for the last Okazaki fragment of the lagging strand is removed, there is no way to polymerase that segment. This leads to ***chromosome***

shortening when the incomplete chromosome replicates again. After each cell division, telomeres continue to shorten, until essential coding information would be lost. In order to prevent this, cells add multiple copies of a simple noncoding sequence to the 3' end of each chromosome. The enzyme that synthesizes ***telomeric DNA repeats*** is called ***telomerase*** (Figure 45.).

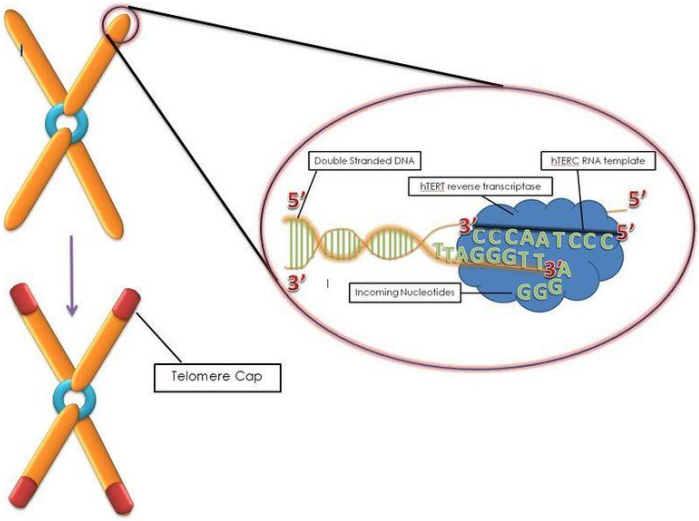


Figure 45. Telomere stabilized and synthesized by telomerase

hTERT, the reverse transcriptase subunit of the telomerase complex, synthesizes the telomeric DNA repeat according to the RNA template of the complex.
https://commons.wikimedia.org/wiki/File:TelomeraseMechanism.jpg
DevelopmentalBiology, CC BY-SA 3.0 <https://creativecommons.org/licenses/by-sa/3.0>, via Wikimedia Commons

3.3 Transcription

Properties and classes of RNA

The major differences between DNA and RNA can be summarized as follows.

1. RNA is a ***single-stranded*** molecule. However, in strict sense only mRNA functions in a single-stranded form. tRNAs and rRNAs form complex 3D shapes by intramolecular base pairing (Figure 46.).
2. The sugar component found in RNA nucleotides is ***ribose***, which has an oxygen-hydrogen pair bound to its 2' carbon atom.
3. Ribonucleotides contain the bases ***adenine, guanine, cytosine, and uracil***.
4. RNA, like enzymes, can catalyze biological reactions. RNA molecules with catalytic activity are called ***ribozymes***.

There are two types of RNAs: "informational" and "functional RNAs".

Informational RNAs code for proteins. They function in the process of decoding genes into polypeptide chains. This class is called ***messenger RNAs (mRNAs)***, because they pass information from DNA to proteins.

The minority of genes code for RNAs, and not proteins, as final products. These RNAs, called ***functional RNAs***, are never translated into polypeptides. ***Transfer RNAs (tRNAs)*** work in the process of translation. They are responsible for bringing amino acids to the mRNA and ribosome. ***Ribosomal RNAs (rRNAs)*** are major components of ribosomes, which are large macromolecular machines, "workbenches" for protein synthesis.

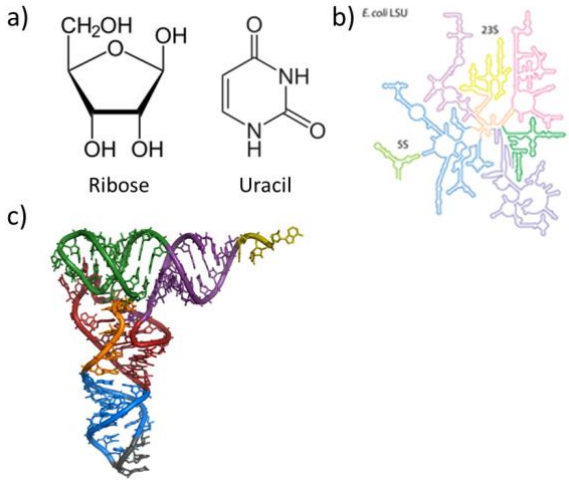


Figure 46. Properties of RNA

a) The sugar moiety in the RNA nucleotides is ribose, and uracil can be found instead of thymine from the nitrogenous bases.
b) 2D representation of the large ribosomal subunit RNA of *E. coli*
c) 3D representation of a yeast tRNA that carries phenylalanine
b) is based on: https://commons.wikimedia.org/wiki/File:T_thermophilus_S_cerevisiae_H_sapiens.png
Petrov AS, Bernier CR, Gulen B, Waterbury CC, Hershkovits E, et al., CC BY-SA 4.0 <https://creativecommons.org/licenses/by-sa/4.0>, via Wikimedia Commons
c) is based on: https://commons.wikimedia.org/wiki/File:TRNA-Phe_yeast_1ehz.png
Yikrazuul, CC BY-SA 3.0 <https://creativecommons.org/licenses/by-sa/3.0>, via Wikimedia Commons

The flow of genetic information

Within the ***DNA*** sequence of any organism's genome, there is information about what gene products the organism can produce, and also information about when, where, and in what amount these products should be produced. However, this information is static (see Figure 3. in Chapter 1, page 6.).

To utilize this information, first an intermediate molecule, a copy of a discrete gene is synthesized. This molecule is ***RNA***, and the process of its synthesis is called ***transcription***. In prokaryotes, the information in RNA is almost immediately utilized to produce a ***polypeptide*** in a process called ***translation***. Transcription and translation take place in the same cellular compartment in prokaryotes. In eukaryotes, transcription and translation are spatially separated. Transcription takes places in the nucleus, and the RNA product undergoes extensive processing before becoming mature mRNA, which is transported into the cytoplasm. Translation occurs in the cytoplasm of the eukaryotic cell.

Overview of the stages of transcription

The first step of information transfer from gene to protein is to ***produce an RNA copy of the nucleotide sequence of a DNA segment*** in a process called transcription. The two DNA strands of the gene to be transcribed separate locally, and one of them acts as a template for RNA synthesis. Both DNA strands can be used as a template, but ***in any one gene, only one strand is used.***

Next, an enzyme called ***RNA polymerase*** synthesizes RNA strand from ***ribonucleotide triphosphates (ATP, GTP, CTP, and UTP)***, following the rules of base complementarity. The ***RNA polymerase works in a 5'- to -3' direction***; in other words, the enzyme adds ribonucleotides always to the 3' growing tip.

Transcription occurs in three steps: ***initiation***, ***elongation*** and ***termination***, which will be described first in prokaryotes, using the gut bacterium *E. coli* as an example.

In most organisms, protein coding genes represent usually short segments of the whole genome. To initiate transcription, the ***RNA polymerase must find the beginning of genes***, the correct starting point for transcription (Figure 47.). The enzyme binds to the promoter, which is a regulatory region of the gene, located close to the start of the transcribed region.

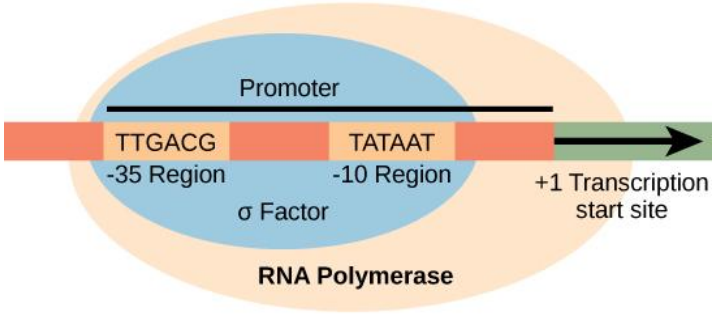


Figure 47. The initiation of transcription

The promoter contains two consensus sequences, the TATA box in the -10 region, and another one in the -35 region. The sigma factor binds to these regions of the promoter, positioning the polymerase correctly.
https://commons.wikimedia.org/wiki/File:Figure_15_02_01.jpg
CNX OpenStax, CC BY 4.0 <https://creativecommons.org/licenses/by/4.0>, via Wikimedia Commons

The bacterial ***RNA polymerase holoenzyme*** is composed of four subunits of the ***core enzyme*** ($2\alpha+\beta+\beta'$) plus a subunit called ***sigma factor*** (σ). Scanning the DNA sequence, the ***sigma factor recognizes the promoter region***, and positions the holoenzyme to initiate transcription at the correct start site. The sigma subunit also has a role in separating the DNA strands so that the core enzyme can bind tightly. Once the core enzyme is tightly bound, it can begin transcription, and the sigma subunit dissociates from the multisubunit complex.

During elongation, the RNA polymerase unwinds the DNA, incorporates ribonucleotides into the growing RNA copy, and rewinds the DNA. A region of single-stranded DNA with a short segment of hybrid nucleic acid can be observed where the RNA polymerase is at work. This is called a ***transcription bubble*** (Figure 48.). RNA chain elongation occurs according to the following general formula:

$$\mathrm{NTP} + (\mathrm{NMP})_n \xrightarrow[\substack{Mg^{2+}\\ \text{RNA}\\ \text{polymerase}}]{\text{DNA}} (\mathrm{NMP})_{n+1} + \mathrm{PP_i}$$

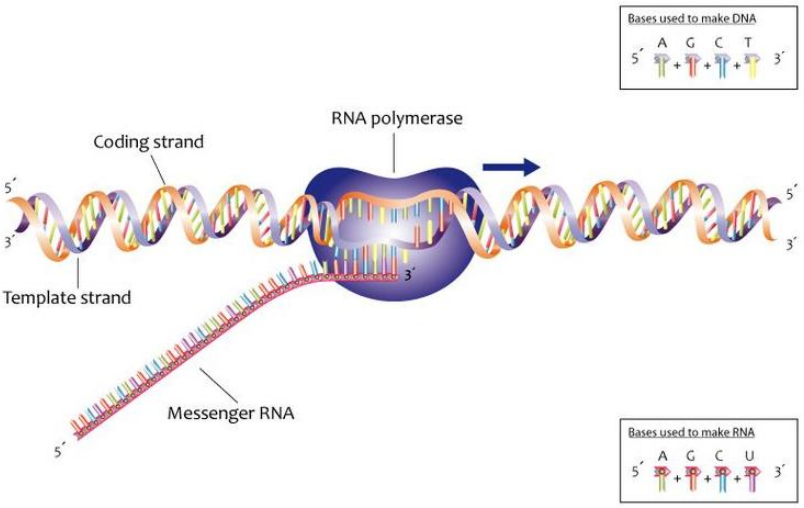


Figure 48. Elongation

The RNA polymerase adds ribonucleotides to the growing RNA strand according to the nucleotide sequence of the template DNA strand. The RNA copy corresponds to the coding DNA strand, except that U stands instead of T in the RNA sequence. Nucleotide databanks store the sequences of the coding strands of genes, in 5'-to-3'direction. As this RNA polymerase transcribe a prokaryotic gene, the primary transcript can be called mRNA.
https://commons.wikimedia.org/wiki/File:Process_of_transcription_(13080846733).jpg
Genomics Education Programme, CC BY 2.0 <https://creativecommons.org/licenses/by/2.0>, via Wikimedia Commons

Elongation continues until RNA polymerase recognizes a nucleotide sequence of the RNA synthesized that acts as a signal for chain termination. This initiates the release of the nascent RNA and the enzyme from the template. When termination occurs according to the intrinsic mechanism, formation of a hairpin loop sets off the release of the completed RNA and the polymerase from the template. The alternative way is to terminate with the help of a protein factor called Rho (Figure 49.).

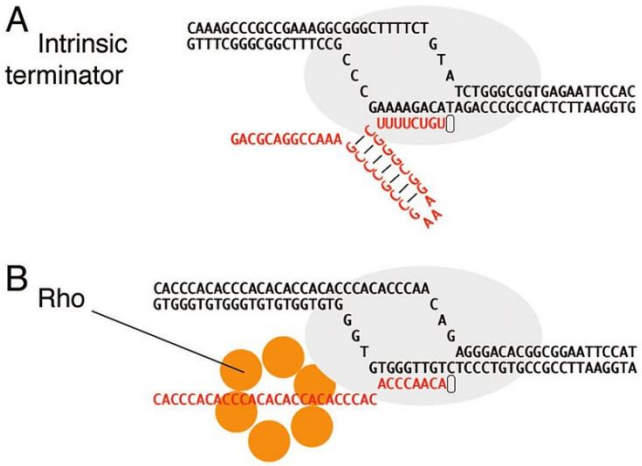


Figure 49. Comparison of intrinsic vs. Rho-dependent termination

https://commons.wikimedia.org/wiki/File:Rho_And_Intrinsic_Termination_photo.jpg
Jeffrey W Roberts, CC BY-SA 4.0 <https://creativecommons.org/licenses/by-sa/4.0>, via Wikimedia Commons

Transcription in eukaryotes

Transcription is more complicated in eukaryotes than in prokaryotes for several reasons.

1. ***Eukaryotic genomes are larger***, and have many ***more genes*** to be transcribed. There is much noncoding DNA in the eukaryotic genomes, so even though eukaryotes have more genes than prokaryotes do, they are far apart from each other. This makes transcription, and especially the initiation step, a much more complicated task.

Eukaryotes divide the job of transcription between three RNA polymerases: ***RNA polymerase I transcribes rRNA genes, RNA polymerase II transcribes the protein-coding genes, and RNA polymerase III makes tRNAs.*** Eukaryotic RNA polymerases consist of more subunits than prokaryotic ones, and they are assisted by several transcription factors.

2. RNA is synthesized in the nucleus, and must be modified before it can be exported out of the nucleus into the cytoplasm for translation. A newly synthesized RNA is called the ***primary transcript or pre-mRNA***. The terms mRNA or mature RNA refer to the product of several modifications that are collectively called ***RNA processing***. The role of RNA pol *II* in these processes will be discussed later.

3. The genomic DNA of eukaryotes is organized into chromatin, whereas that of the prokaryotes is virtually naked. ***Changing the chromatin structure*** has a very sophisticated role in the regulation of gene expression in eukaryotes.

Three steps of pre-mRNA processing in eukaryotes

RNA processing in eukaryotes include (1) the addition of a ***cap*** at the ***5' end***, (2) the addition of a ***3' tail of adenine nucleotides***, and (3) ***splicing to eliminate introns*** (Figure 50.). The protein tail of ***RNA polymerase II***, called the ***carboxyl tail domain (CTD)***, plays a central role in coordinating all processing events (Figure 51.).

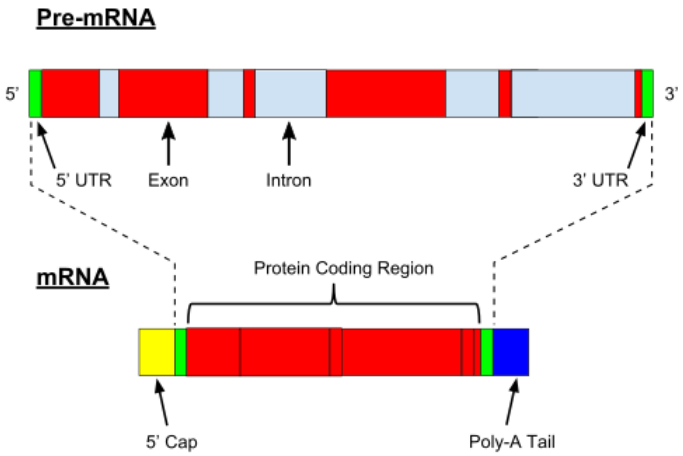


Figure 50. Differences of the primary transcript (pre-mRNA) and the mature mRNA

UTR: untranslated region; for further description, refer to the main text.
https://commons.wikimedia.org/wiki/File:Pre-mRNA.svg
Nastypatty, CC BY-SA 4.0 <https://creativecommons.org/licenses/by-sa/4.0>, via Wikimedia Commons

When the nascent RNA first emerges from RNA pol II, ***a specially-bound nucleotide, called the cap*** is added to its 5' end. The cap protects the mRNA from degradation, and it is required for translation.

RNA synthesis continues until the RNA polymerase meets a sequence called polyadenylation signal. Then an enzyme cuts off the RNA, and to this cut end a stretch of 150 to 200 adenine nucleotides called the ***poly(A)tail*** is added.

The majority of eukaryotic genes contain ***introns***, segments of DNA that ***do not code for polypeptides***. Introns are removed from the primary transcript while the RNA is still being synthesized. The CTD of polymerase *II* serves as a binding site for the enzymes of the RNA splicing machinery. The spliceosome, a molecular machine made of small nuclear ribonucleoprotein particles, catalyzes the removal of introns. The removal of introns and joining of exons is called ***splicing***. The mature mRNA contains only ***exons***, the ***protein-coding regions*** of the gene.

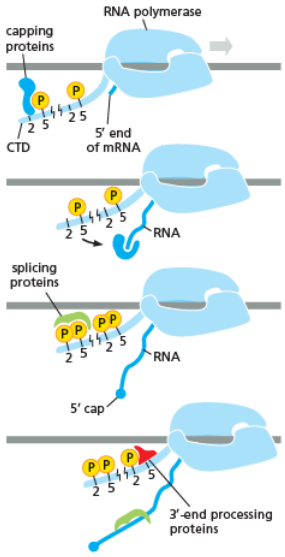


Figure 51. The role of polymerase II CTD in RNA processing

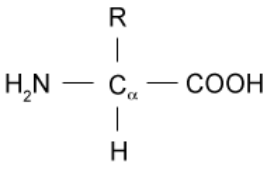
Reversible phosphorylation of the amino acids of the CTD creates binding sites for the enzymes involved in RNA processing
https://commons.wikimedia.org/wiki/File:Ctd_role_.png
Ravinesh rds, CC BY-SA 4.0 <https://creativecommons.org/licenses/by-sa/4.0>, via Wikimedia Commons

3.4 Translation

Protein structure

The final stage of information transfer is the translation. ***Translation means the synthesis of a polypeptide directed by an RNA sequence.*** Before discussing the details of this process, we need to recall protein structure.

Proteins are polymer macromolecules composed of ***amino acid*** monomers. All amino acids have an amino group ($-NH_2$), a carboxyl group (-COOH), and a variable side chain, the R group.



Proteins synthesized by living organisms are built up of 20 natural amino acids. The amino acids are linked together by covalent bonds called ***peptide bonds*** (Figure 52.).

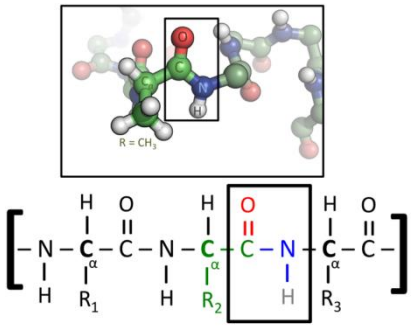


Figure 52. Amino acids linked by peptide bonds

Three amino acids as the portion of a polypeptide chain is presented. The four atoms that form the peptide group are highlighted.
https://commons.wikimedia.org/wiki/File:Peptide-Figure-Revised.png
Chemistry-grad-student, CC BY-SA 3.0 <https://creativecommons.org/licenses/by-sa/3.0>, via Wikimedia Commons

Proteins have a complex structure that has four levels of organization (Figure 53.). The ***primary structure*** of a protein refers to the ***sequence of amino acids*** in the polypeptide chain. The ***secondary structure*** refers to regular sub-structures of the polypeptide chain. The most common secondary structures are the ***α-helix*** and the ***β-strand or β-pleated sheet***. These secondary structures are defined by hydrogen bonds between the main-chain peptide groups. The ***tertiary structure*** refers to the ***three-dimensional structure*** of a protein, produced by the folding of the secondary structures. Many proteins are ***globular***, like e.g. enzymes and antibodies. Proteins with linear shape are called ***fibrous*** proteins. Many of them fulfill structural role, like e.g. in hair or muscles. Proteins that are ***composed of two or more polypeptide chains***, also called subunits, have a ***quaternary structure***.

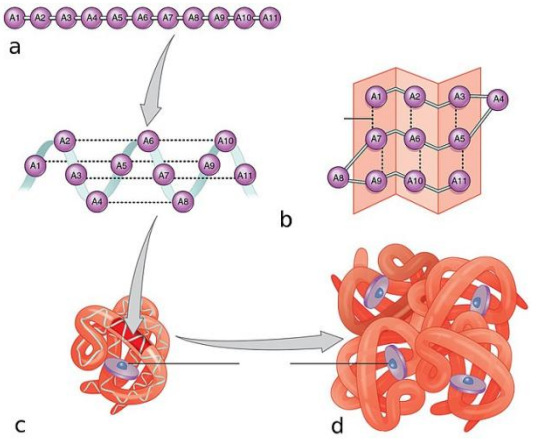


Figure 53. Four levels of protein structure

a) Primary structure; b) Secondary structures: α-helix (left) and β-sheet (right); c) Tertiary structure of a β-globin chain.The heme group (purple) is a non-protein ring structure with an iron atom at its center; d) Hemoglobin is an example of heterotetromer, composed of two α and two β subunits.
https://commons.wikimedia.org/wiki/File:225_Peptide_Bond-01_labeled.jpg
OpenStax, CC BY 3.0 <https://creativecommons.org/licenses/by/3.0>, via Wikimedia Commons

The "protagonists" of translation

The information for assembling a polypeptide chain arrives to the translation machinery in the form of an ***mRNA***. ***The linear sequence of nucleotides in a gene determines the linear sequence of amino acids in a protein. The genetic code is the set of rules that specifies how this is done*** (Figure 54.).

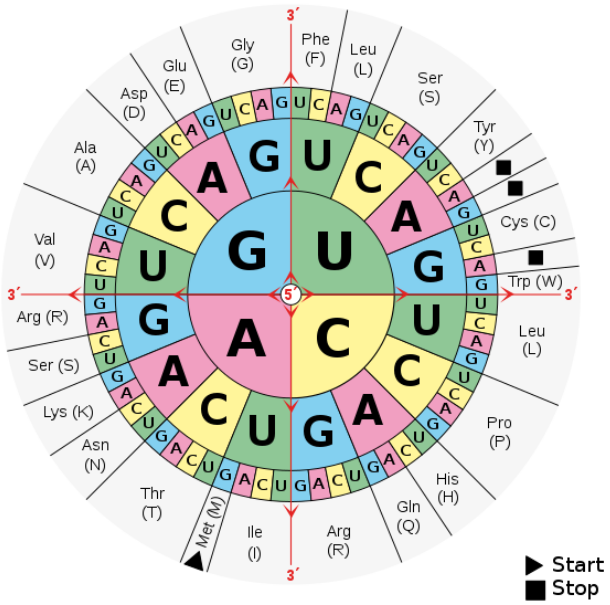


Figure 54. The genetic code

https://commons.wikimedia.org/wiki/File:Aminoacids_table.svg; Mouagip, Public domain, via Wikimedia Commons

Three nucleotides encode an amino acid. These ***triplets*** are termed ***codons***. The genetic code contains altogether 64 codons, of which 61 specify amino acids, and three are stop codons. Stop codons are signs for the termination of translation. The code is read from a fixed starting point in a ***non-overlapping*** fashion. The genetic code is ***degenerate*** (also called ***redundant***), which means that some amino acids are specified by more than one codon. The genetic code is ***universal***, that is, virtually all organisms on Earth use this same genetic code. (A few differences in codon usage have been observed in mitochondrial genomes.)

Each amino acid is attached to a specific ***tRNA*** (Figure 55.), which brings that amino acid to the ribosome, the site for protein synthesis. The "flattened" 2D structure of a tRNA is similar to a cloverleaf. The amino acid is attached to the "stem" of the cloverleaf (the 3' end of the tRNA) by a special enzyme called ***aminoacyl-tRNA synthetase***. The middle "leaflet" of the cloverleaf (loop of tRNA) is called the ***anticodon loop*** because it carries a nucleotide triplet called anticodon. Each amino acid has a specific synthetase that links it only to those tRNAs that recognize the codons for that particular amino acid.
Several different tRNAs may be charged with the same amino acid. These are called ***isoaccepting tRNAs***. Certain tRNAs recognize and bind to several alternative codons, not just one with complementary sequence. This is possible because of a loose kind of pairing between the 3' end of the codon and 5' end of the anticodon. This loose pairing is called ***wobble***.

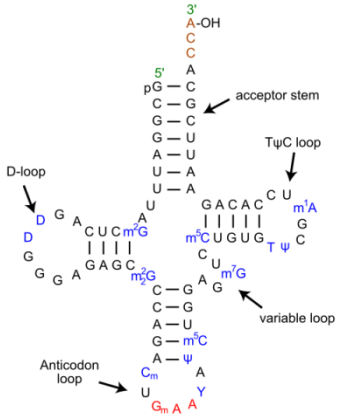


Figure 55. tRNA for carrying phenylalanine in yeast

tRNAs often have modified bases shown in blue; The anticodon loop and the site for attaching the amino acid are shown in red.
https://commons.wikimedia.org/wiki/File:TRNA-Phe_yeast_en.svg
Yikrazuul, CC BY-SA 3.0 <https://creativecommons.org/licenses/by-sa/3.0>, via Wikimedia Commons

For protein synthesis to take place, a charged tRNA and an mRNA molecule must associate with a ***ribosome*** (Figure 56.). The task of the tRNA and the ribosome is to translate the nucleotide sequence of the mRNA into the amino acid sequence of a protein.

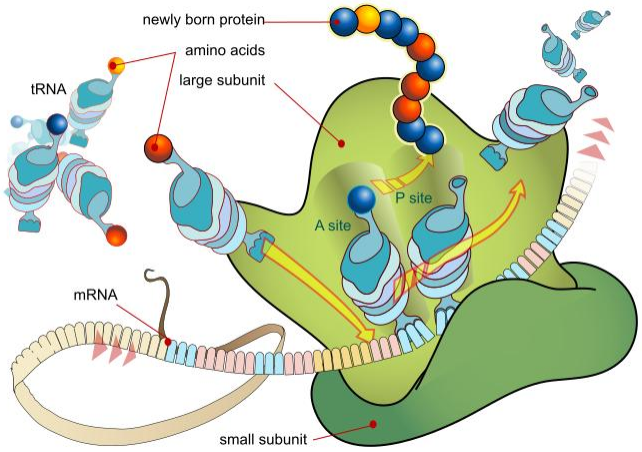


Figure 56. Ribosome at work

This somewhat artistic diagram illustrates how a ribosome, an mRNA, and lots of tRNA molecules work together to produce a polypeptide. For a more accurate description of the structure and process, refer to the main text.
https://commons.wikimedia.org/wiki/File:Ribosome_mRNA_translation_en.svg
LadyofHats, Public domain, via Wikimedia Commons

Ribosomes consist of a small and a large subunit; both of them ***made up of ribosomal RNAs (rRNAs) and proteins.*** In ***prokaryotes***, the small and the large subunits are called ***30S*** and ***50S***, respectively. They associate to form a ***70S*** particle. (S stands for Svedberg, the unit for sedimentation coefficient.) ***Eukaryotic ribosomes*** are built up of ***40S*** small, and ***60S*** large subunits, forming ***80S*** ribosomes when associate.
The goal is to position the tRNA and mRNA molecules on the ribosome so that the codon of the mRNA can interact with the anticodon of the tRNA.

The binding site for mRNA is completely within the small subunit. There are three binding sites for tRNA molecules. The ***A site*** binds an incoming charged tRNA whose anticodon is complementary to the codon in the A site of the small subunit. The tRNA in the ***P site*** contains the growing peptide chain. The ***E site*** contains a tRNA that no longer carries an amino acid, and is ready to leave the ribosome.
The ***decoding center*** in the 30S subunit ensures that only tRNAs providing a correct anticodon-codon match are accepted at the A site. The ***peptidyl transferase center*** in the 50S subunit is the place where peptide-bond formation is catalyzed. Both centers are composed entirely of rRNA. ***Ribozymes work in the ribosomes***.

Stages of translation

Translation, similarly to the other processes of information transfer, will be discussed in division for three stages: initiation, elongation, and termination (Figure 57.).

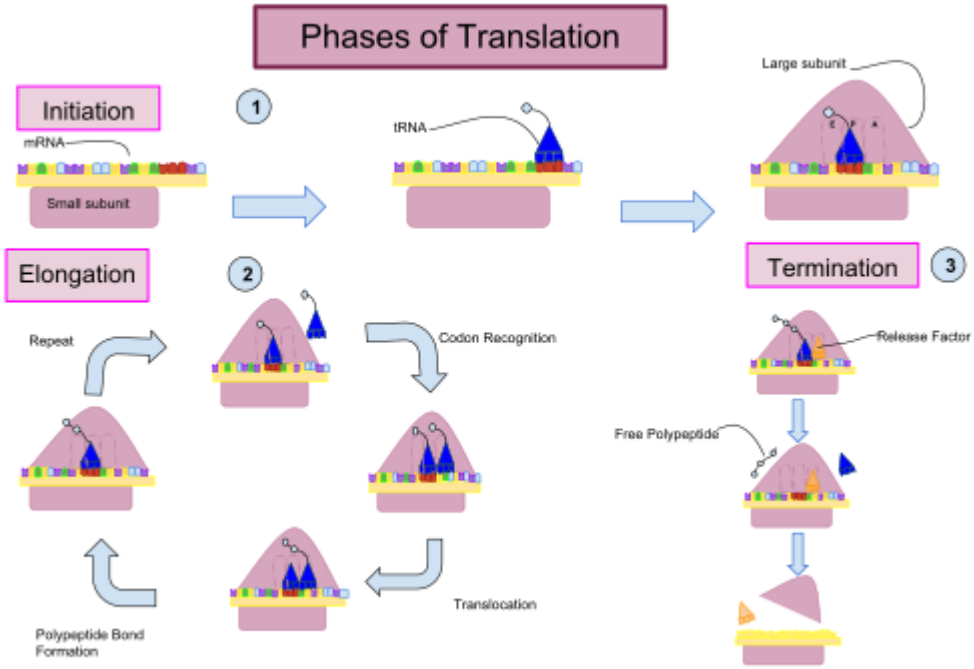


Figure 57. Simple demonstration of the phases of translation

For a detailed description of the stages, refer to the main text.
https://commons.wikimedia.org/wiki/File:Translation_drawing-_Carina_Huerta.svg
Carihuer, CC BY-SA 4.0 <https://creativecommons.org/licenses/by-sa/4.0>, via Wikimedia Commons

The main task of ***initiation*** is to place the ***first aminoacyl-tRNA in the P site*** of the ribosome and, in this way, establish the correct reading frame of the mRNA. The first amino acid in any newly synthesized polypeptide is ***methionine*** (fMet in prokaryotes), ***specified by the codon AUG***, and inserted by a special initiator tRNA.
In prokaryotes, matching a special sequence of the rRNA in the small ribosomal subunit with its complementary sequence in the mRNA helps to assemble the 30S initiation complex.
In eukaryotes, the 40S ribosomal subunit assisted by other factors recognizes the cap of the mRNA, and scans for an AUG codon where translation can begin.

Both in prokaryotes and eukaryotes, initiation factors assist the assembly of the ribosome at the start site, which is completed by joining of the large subunit.

The steps of ***elongation*** are as follows. ***Charged tRNAs*** bringing the amino acid specified by the next codon ***bind to the A site*** in the ribosome. In the peptidyl transferase center, a ***peptide bond is formed*** between the last amino acid of the growing polypeptide chain and the incoming amino acid, while the whole peptide chain is transferred onto the newly arrived tRNA. A ***translocation step*** positions the empty tRNA in the E site and the tRNA carrying the polypeptide chain in the P site, respectively. The ***empty tRNA leaves*** the E site, and the A site is ready to receive the next charged tRNA. During the elongation process, two protein factors support the work of the ribosome.

The steps of elongation cycle continues until the ribosome meets a ***stop codon*** in the A site. Stop codons are signs for ***termination***. Stop codons are recognized not by tRNAs, but proteins called ***release factors***. The interaction bewteen the release factors and the A site leads to the termination of the elongation process, and the disassembly of the ribosome.

When released from the ribosome, newly synthesized proteins must fold into their correct three-dimensional shape to achieve their native conformation. Many proteins are modified by the covalent attachment of different molecules to the R group of certain amino acids.

4. Changes of the Hereditary Material

4.1 Mutation

Basic concepts

Every classical genetic analysis begins with the selection of ***variants*** – these are organisms that differ in some particular character. How do these genetic variants arise? The ***origin of the genetic diversity*** in a population is a process called mutation. ***Mutation is a heritable change in the genetic material***, an alteration in the genome of an organism.

Mutation can take place at two different levels. In ***gene mutation***, an allele (variant) of a gene changes, becoming a different allele. Because this is a small-scale change in the genetic material, often affecting only a single or some nucleotides, a gene mutation is also called ***point mutation***. When ***chromosome mutation*** occurs, segments of chromosomes, whole chromosomes, or even entire sets of chromosomes change. These events will be discussed later, in the section entitled "Large-scale chromosome changes".

The ***wild-type*** is a reference point or standard in genetics. The wild-type allele means the normal, natural form found in nature, or the standard laboratory stock (e.g. wild-type and "vestigial" mutant flies are shown in Figure 4, page 6). When the wild-type allele changes into any other form, the process is a ***forward mutation***. Any change back to the wild-type is called ***reverse mutation*** or ***reversion*** or ***back mutation***. The non-wild-type allele of a gene is also called a "mutation", so in practice both the process and the result of the process are called by this name.

$$\left.\begin{matrix} a^+ \rightarrow a \\ D^+ \rightarrow D \end{matrix}\right\} \text{ forward} \qquad \left.\begin{matrix} a \rightarrow a^+ \\ D \rightarrow D^+ \end{matrix}\right\} \text{ reverse}$$

Other useful terms are: ***mutant, mutation event, mutation frequency***. A mutant is an organism or a cell whose changed phenotype is attributable to the possession of a mutation. A mutation event is the actual occurrence of a mutation. The mutation frequency shows the proportion of mutations in a population of cells or individuals.

Germinal versus somatic mutation

Germinal mutations occur in the ***germ line***, a special tissue set aside during development to produce gametes. Germinal mutations ***may be transmitted*** to some or all progeny. An individual of normal phenotype can harbor undetected mutant gametes. If a mutant gamete participates in fertilization, the mutation will be passed onto the next generation, and the phenotype caused by the mutation can be detected in the progeny. Genetic diversity in the population is attributable to germ line mutations.

Somatic mutations occur in the cells of ***somatic tissues***. Somatic mutations are ***not transmitted to progeny***. When a mutated somatic cell divides repeatedly, it forms a population of identical cells, called a clone. Staying together during development, these cells can produce an observable outcome of the somatic mutation, called a mutant sector. The earlier the mutation event in development takes place, the larger the mutant sector will be (Figure 58.).

Somatic mutations do not usually cause serious problems in an adult, unless they affect specific genes, such as ***proto-oncogenes*** or ***tumor suppressor genes***. These genes are regulators of the cell cycle and programmed cell death; therefore, mutations in these genes can lead to severe consequences, such as cancer.

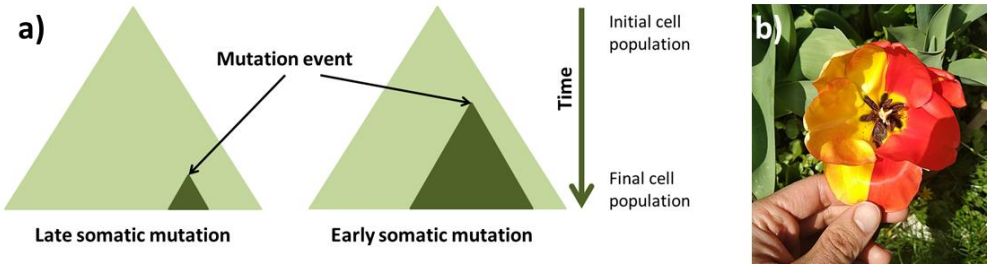


Figure 58. Somatic mutation

a) The earlier the somatic mutation, the larger the mutant sector. b) Mosaicism in a tulip flower. Mosaicism can result from somatic mutation which, in this case, must have happened very early during the development of this flower.
b) Based on: https://commons.wikimedia.org/wiki/File:Tulip_with_mosaicism.jpg
Paddlestroke, CC BY-SA 4.0 <https://creativecommons.org/licenses/by-sa/4.0>, via Wikimedia Commons

Mutant types

Morphological mutations affect the outwardly visible properties of an organism, such as ***shape, color, or size***. For example, the eye and wing mutations of *Drosophila* (Figure 59.), Mendel's garden pea variants, or the mammal fur color variants are morphological mutations. Color mutations are valuable properties in plant breeding. Mutations in the chloroplast genes result in mosaic or striped color variation of the leaves (Figure 59.).

Figure 59. Morphological mutants

a) Drosophila eye colors (clockwise): brown, cinnabar, sepia, vermilion, white, wild-type. The white-eyed fly has a yellow body, the sepia-eyed fly has a black body, and the brown-eyed fly has an ebony body.
b) Variegated ivy.
a): https://commons.wikimedia.org/wiki/File:EyeColors.jpg
No machine-readable author provided. Ktbn assumed (based on copyright claims)., Public domain, via Wikimedia Commons
b): Author: Heron; https://creativecommons.org/licenses/by-sa/3.0/

Lethal mutant alleles can be recognized by their effect on the survival of an organism. Lethal mutations ***reduce the viability of the individual*** (Figure 60.). In genetics, any mutation that interrupts the development of an individual before it can reproduce is considered lethal. Genes – whose loss-of-function mutations are lethal – are called essential genes.

Figure 60. Manx cat

The tailless Manx cats carry a recessive mutant allele that is lethal in homozygous form.
https://commons.wikimedia.org/wiki/File:Manx_cat_by_Karen_Weaver.jpg
Karen Weaver, CC BY 2.5 <https://creativecommons.org/licenses/by/2.5>, via Wikimedia Commons

When an individual carries a ***conditional mutation***, the mutant allele causes a mutant phenotype in only a certain environment. This is called the ***restrictive condition***. In some different environment, called the ***permissive condition***, the individual shows wild-type phenotype. Geneticists often use breeding temperature as a restrictive/permissive condition. Mutants sensitive to the breeding temperature are called ***temperature-sensitive*** (Figure 61.).

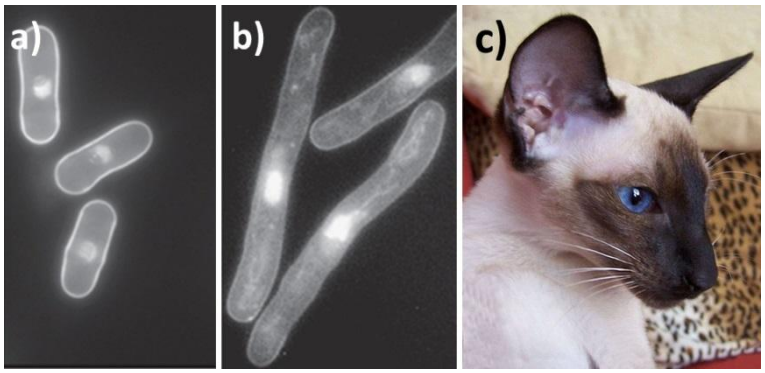


Figure 61. Temperature-sensitive mutants

Wild-type cells (a) and temperature sensitive mutants (b) of the fission yeast, *Schizosaccharomyces pombe*. Temperature-sensitive mutants of yeasts have an outstanding role in cell cycle research. c) Siamese cat. The fur color is the result of a temperature sensitive mutation in the "*C*" gene.
https://commons.wikimedia.org/wiki/File:Modern_Seal_Point_Female.jpg
Thaifong, CC BY-SA 3.0 <https://creativecommons.org/licenses/by-sa/3.0>, via Wikimedia Commons

Biochemical mutations are identified by the ***loss or change of some biochemical function***. Biochemical mutations have been extensively analyzed in microorganisms. Microorganisms that can exist on a minimal medium containing simple inorganic salts and an energy source are called ***prototrophic***. Biochemical mutants, however, are often ***auxotrophic.*** They must be supplied with certain additional nutrients which they are unable to synthesize themselves (Figure 62.). Many human hereditary diseases are biochemical mutations. The term ***"inborn errors of metabolism"*** has been used to describe such biochemical disorders.

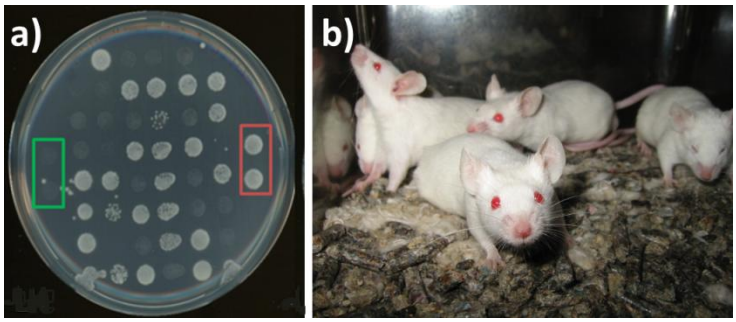


Figure 62. Biochemical mutants

a) Testing fission yeast strains for nutrient requirement. Auxotrophic mutants (green box; control) are unable to grow on minimal medium, whereas fully prototrophic strains (red box; control) grow well.
b) Albino mice. Albinism in mammals is usually due to mutation(s) in a pigment synthesis pathway.
https://commons.wikimedia.org/wiki/File:Lightmatter_lab_mice.jpg
by Aaron Logan, CC BY 1.0 <https://creativecommons.org/licenses/by/1.0>, via Wikimedia Commons

Classification of mutations according to their effect on gene function

Mutation events are generally destructive, and delete or change crucial functional regions of the gene. Such a change interferes with wild-type functions, and the result is a ***loss-of-function mutation***. Generally loss-of-function mutations are found to be recessive. When the loss of function is complete, a ***null mutation*** is generated (Figure 63.).

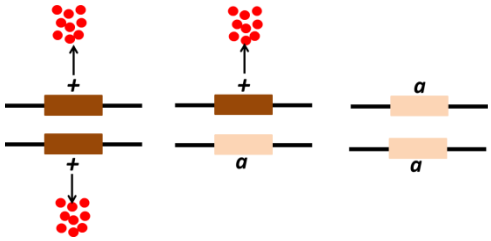


Figure 63. Null mutation

Mutant allele *a* has completely lost its function. However, the product of the wild-type allele in the heterozygote is often enough to produce wild-type phenotype.

When the inactivation is incomplete, the new allele is said to be a ***leaky mutation*** (Figure 64.).

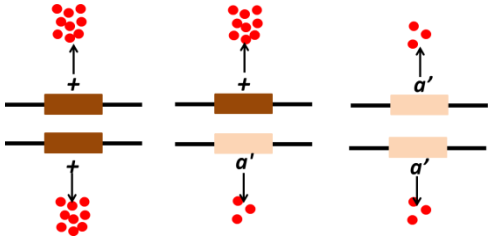


Figure 64. Leaky mutation

Mutation *a'* still retains some function, but the product of the mutant allele in the homozygote is not enough to produce a wild-type phenotype.

Mutational events introduce random genetic changes; in most cases, they result in the disruption of gene function, and thus the production of a loss-of-function mutation. Sometimes the random change by pure chance confers some new function on the gene. It may participate in a new reaction, or increase its activity. However, ***gain-of-function*** mutations are less common (Figure 65.). In a heterozygote the new function will be expressed, and therefore the gain-of-function mutations are ***dominant***. Gain-of-function mutations in homozygous form are often lethal.

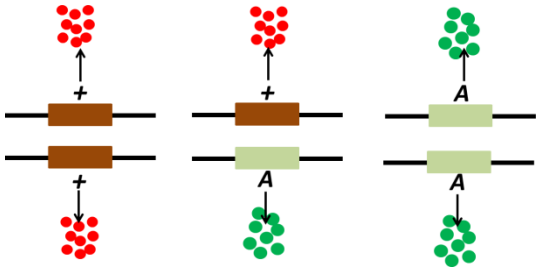


Figure 65. Gain-of-function mutation

Mutation *A* has acquired a new cellular function. *A* in the heterozygote is most likely dominant. The homozygous mutant may or may not be viable, depending on the role of the wild-type allele.

Point mutations at the molecular level

At DNA level, point mutations are classified into two main types: base ***substitutions*** and base ***additions or deletions*** (collectively termed ***indel mutations*** for insertion and deletion.)

1. ***Base substitutions*** can also be divided into two subtypes: transitions and transversions.

a) ***Transition*** occurs when a base is replaced by another base of the same chemical type: a purine base is replaced by a purine base, or a pyrimidine base is replaced by a pyrimidine base, e.g.: A→G or T→C.

b) ***Transversion*** means the replacement of a base with a base of different chemical type, that is, a purine base is replaced by a pyrimidine base or a pyrimidine base is replaced by a purine base, e.g.: A→T or C→G.

2. ***Addition or deletion*** mutations are actually addition or deletion of nucleotide pairs. In the simplest case, these mutations are single-base-pair additions or single-base-pair deletions.

Point mutations may have different consequences ***at protein level***.

Synonymous or silent mutations change one codon for an amino acid into another codon for that same amino acid.

Missense mutations change the codon for one amino acid into a codon for another amino acid. In case of ***conservative missense mutations***, the new codon specifies a chemically similar amino acid. Such a change does not alter protein function generally. In case of ***nonconservative missense mutations***, the new codon specifies a chemically dissimilar amino acid. This can have a severe effect on protein function.

Nonsense mutations change the codon for one amino acid into a translation termination (stop) codon. This leads to truncation of the encoded protein.

The addition or deletion of a single base pair changes the reading frame for the remainder of the translation process, therefore these lesions are called ***frameshift mutations***. Frameshift mutations typically result in complete loss of protein structure and function (Figure 66.).

The fat cat ate the big rat.
The xfa tca tat eth ebi gra t.

Figure 66. Frameshift mutation

A playful explanation for how frameshift mutations can cause severe problems.

The origin of mutations

Newly arising mutations are categorized as ***induced*** or ***spontaneous*** mutations. Induced mutations are produced intentionally; they are resulted from treatment with ***mutagens***, agents that are known to increase the rate of mutations. Exposure to mutagens is called ***mutagenesis***, and the resulted organism is said to be ***mutagenized***. The most commonly used mutagens are high-energy radiations (***physical mutagens***) and specific chemicals (***chemical mutagens***). When point mutations are induced, often a linear dose response is observed: the greater the dose of mutagen, the greater the number of mutation induced.

Spontaneous mutations arise ***in the absence of known mutagen treatment***. They account for the background rate of mutation, the ultimate source of genetic variation that is seen in genetic populations. The frequency of induced mutations is significantly higher than that of spontaneous mutations.

Spontaneous mutations

Spontaneous mutations arise from a variety of sources, including errors in DNA replication, spontaneous lesions, and the insertion of transposable elements.

Errors in DNA replication:

1. Changes of bases from one ***tautomeric form*** to another can ***result in mispairing***. The rare imino and enol forms of bases lead to irregular base pairing (Figure 67.). During replication, this can lead to ***base substitution***; however, polymerase *III* can repair the majority of such mistakes.

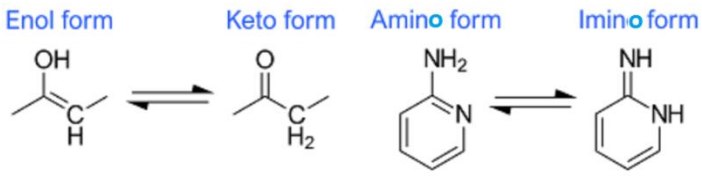


Figure 67. Tautomeric forms

Bases are usually in the keto and amino form, and pair regularly. When these chemical groups rearrange, the rare imino forms of cytosine and adenine make them capable of pairing irregularly. Similarly, in the rare enol form, thymine and guanine can pair with each other.

2. ***Replication of repetitive sequences*** may lead to ***frameshift***. Repetitive sequences may lead to replication slippage.When the newly synthesized strand “slips”, this results in base addition. Slippage of the template strand results in deletion (Figure 68.). When indel mutations add or subtract a number of bases not divisible by three, they produce frameshift mutations in the protein-coding regions.

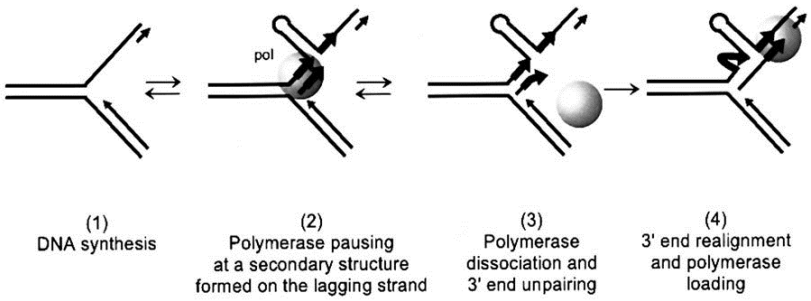


Figure 68. Polymerase slippage

In this case, a secondary structure in the lagging strand caused the polymerase to slip. As a result, a portion of the template strand is deleted.

Spontaneous lesions:

1. Spontaneous lesions are naturally occurring damages to DNA. Two of the most frequent ones are depurination and deamination, the former being the more common.

Depurination means the loss of a purine base (Figure 69.). During replication, the apurinic site cannot specify any kind of base.

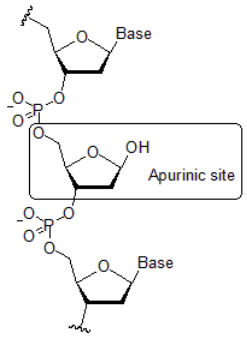


Figure 69. Apurinic site

2. The ***deamination*** of cytosine yields uracil. Unless corrected, uracil pairs with adenine in the course of replication, resulting GC → AT transition. The deamination of 5-methylcytosine generates thymine, which is a naturally occurring base in DNA. Thus, deamination of 5-methylcytosine sites frequently produce C to T transitions (Figure 70.).

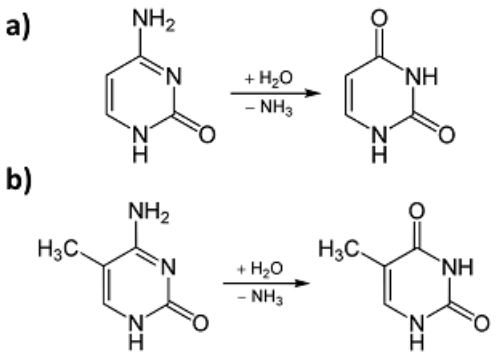


Figure 70. Deamination of bases

a) Deamination of cytosine to uracil; b) Deamination of 5-methylcytosine to thymine.

3. Normal aerobic metabolism produces different active oxygen species as by-products. Reactive oxygen species, like superoxide radicals ($O_2^{\cdot-}$), hydrogen peroxide (H_2O_2), and hydroxyl radicals (·OH) cause ***oxidative damage to DNA and nucleotides***, resulting in mutation (Figure 71.).

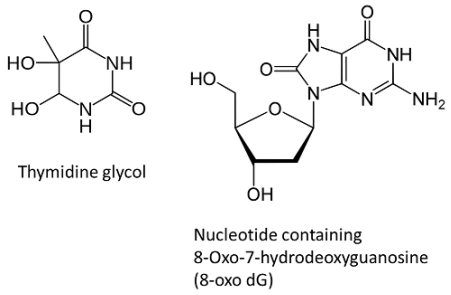


Figure 71. Oxidatively damaged bases

Transposable elements:

Genetic studies of maize started in the 1930s yielded results that changed the classical picture of genes and chromosomes (Figure 72.). Some findings were possible to interpret only if scientists assumed that ***certain genetic elements can move from one location to another in the genome***. A variety of names have been used to call these elements, e.g.: ***controlling elements, jumping genes, mobile genes, transposable elements, transposons.***

Transposable elements can move to new positions within the same chromosome or even to different chromosomes. They can be detected through the mutations that they produce when they inactivate genes into which they insert. Because of this feature, we call them ***"biological mutagens"***.

Transposable elements have been detected in model organisms, as well as in a variety of microbes, plants, and animals. ***They exist in virtually all organisms***. Despite their abundance, their role is not clear yet.

Figure 72. Activity of a transposable element causes mosaicism in maize

Insertion of a transposable element into a gene responsible for pigment production disrupts pigment production, resulting in yellow kernels. Excision of the transposable element restores the activity of the gene, and so the pigment production. The larger the spots on the kernel are, the earlier the excision occurred during its development.

https://commons.wikimedia.org/wiki/File:PLoS_Mu_transposon_in_maize.jpg

Induced mutations

Mutagens act at least by three different mechanisms. They can be incorporated into the DNA to ***replace*** a base; they can ***alter*** a base, causing specific mispairing; they can ***damage*** a base so that it cannot pair with any base. Each mutagen can be characterized by a distinct ***mutational specificity***, that is, each has a preference for a certain type of mutation.

Chemical mutagens:

1. ***Base analogs***, like e.g. ***5-bromouracil (5-BU) and 2-aminopurin (2-AP)***, are chemical compounds similar to the normal nitrogenous bases of DNA (Figure 73.). They can incorporate into DNA occasionally in place of normal bases. They are easily ionized and change to rare tautomeric forms (enol, imino). This results in mispairing and base substitution in the subsequent replication.

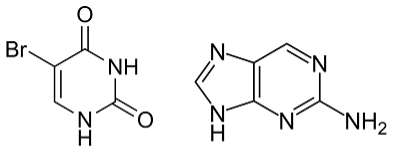


Figure 73. Base analogs

2. ***Alkylating agents*** are commonly used as mutagens. They alter a base, thereby causing mispairing. ***Ethylmethanesulphonate (EMS)*** adds an ethyl group, while ***nitrosoguanidine (NG)*** adds a methyl group to different positions on all four bases (Figure 74.).

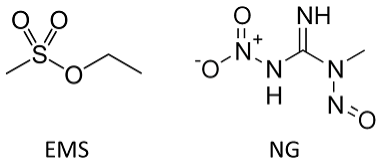


Figure 74. Alkylating agents

3. ***Intercalating agents*** are flat, planar molecules, consisting of three ring structures. They mimic base pairs, and are able to slip themselves in (intercalate) between the nitrogenous bases of DNA (Figure 75.). They can cause single-nucleotide-pair insertions or deletions.This group of compounds includes ***proflavin, acridine orange***, and the ***ICR compounds***.

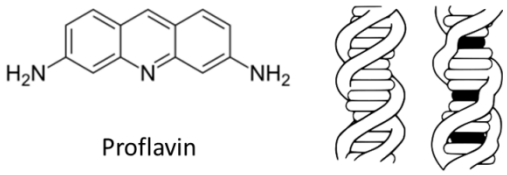


Figure 75. Intercalating agent and its effect on DNA

https://commons.wikimedia.org/wiki/File:DNA_intercalation.svg
Gigillo83, Public domain, via Wikimedia Commons

Physical mutagens:
Ultraviolet light generates a number of distinct types of alterations in DNA, called photoproducts. The most typical ones are the ***cyclobutane pyrimidine photodimer*** and the ***6-4 photoproduct*** (Figure 76.). Both lesions arise by linking adjacent pyrimidine bases in the same strand. The UV photoproducts perturb the local structure of DNA, and lead to the blockage of DNA replication.
Ionizing radiations severely damage DNA, both directly and indirectly. They create reactive oxygen species in the cell, depurinate the DNA, and cause ***double-stranded DNA breaks***. This latter is the main cause of the lethal effect of ionizing radiations.

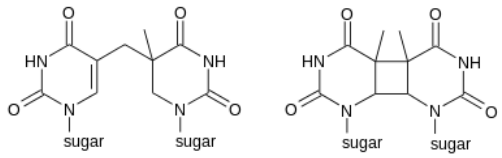


Figure 76. 6-4 photoproduct and cyclobutane pyrimidine photodimer

4.2 Recombination

Basic concepts

The ***production of new allele combinations*** is called recombination. Recombination is a natural part and distinguished feature of meiosis. ***At meiosis, recombination generates haploid meiotic products, whose genotypes are different from the genotypes of those haploid cells that united to form the meiocyte.***

There are two mechanisms for recombination: independent assortment and crossover. Recombination of ***unlinked genes*** (genes that reside on separate chromosomes) occurs by ***independent assortment***. Recombination of ***linked genes*** (genes that are located on the same chromosome) occurs by ***crossing over***. Recombination between linked genes can be used to define their distance apart on the chromosomes. ***Genetic maps*** (also called linkage maps) show the position of genes on the chromosomes (for an example, see Figure 12. on page 10).

Recombination by independent assortment

Recombination by independent assortment is also called ***interchromosomal recombination***, because intact parental chromosomes are re-distributed to the meiotic products in this process. ***Recombinants*** (meiotic products with new allele combinations) are easy to detect in simple organisms (Figure 77.).

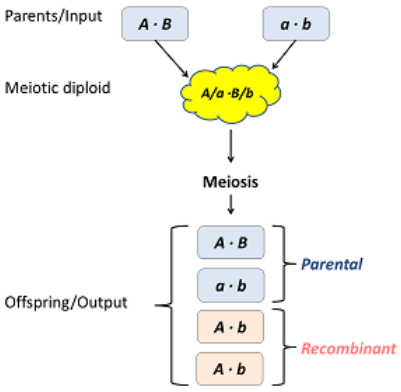


Figure 77. Recombination in organisms with haploid life cycle

Detection of recombinants is straightforward in organisms with haploid life cycle such as fungi or algae, because the input and output genotypes can directly be inferred from the phenotype of the parental and offspring individuals.
However, detection of recombinants in plants and animals, organisms that have a diploid life cycle, requires a specific experimental method (Figure 78.).

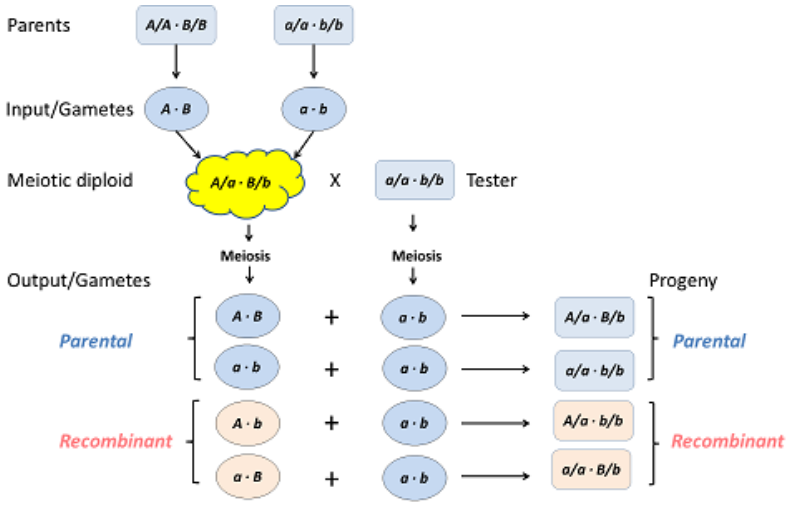


Figure 78. Recombination in organisms with diploid life cycle

In organisms with diploid life cycle, the input and output cells of meiosis are gametes, whose genotypes can not be detected directly. To detect recombinant output gametes, we must perform a ***testcross.*** In testcross, the heterozygous diploid individual is crossed with a recessive tester. In this case, the genotype of the gametes produced by the heterozygote can be inferred from the phenotypes of its offspring.
Independent assortment of chromosomes produces a recombination frequency of 50% (Figure 79.).

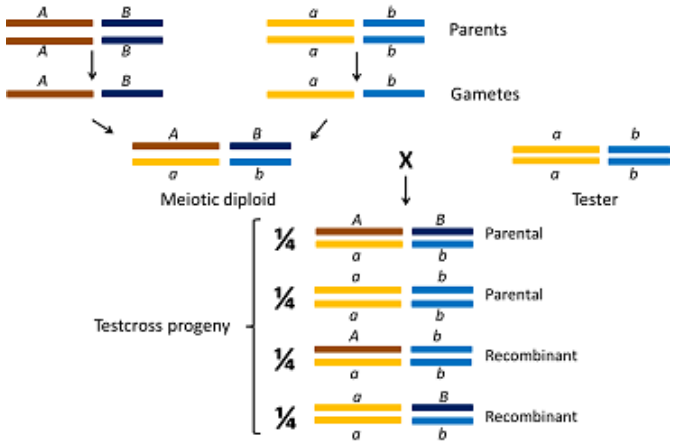


Figure 79. Independent assortment of chromosomes

Two chromosome pairs of a diploid organism, with "*A*" and "*a*" alleles on one chromosome, and "*B*" and "*b*" alleles on the other chromosome are shown.

Recombination by crossing over

When two genes are localized close together on the same chromosome (linked), they can produce new allele combinations only by crossing over. This type of recombination is also called ***intrachromosomal recombination***. Simply interpreted, ***cossing over is a breakage-and-rejoining process***. Two DNA molecules break at the same position between two gene loci, and rejoin in two reciprocal nonparental combinations (Figure 80.). Cytologists can observe crossover sites as chiasmata.

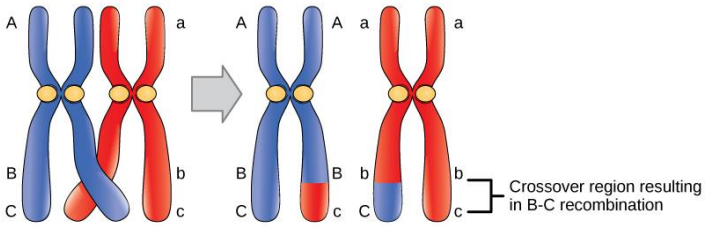


Figure 80. Simple presentation of a crossing over

Based on: https://commons.wikimedia.org/wiki/File:Figure_17_02_01.jpg
CNX OpenStax, CC BY 4.0 <https://creativecommons.org/licenses/by/4.0>, via Wikimedia Commons

Crossing over occurs between non-sister chromatids. When it occurs, half of the meiotic products is recombinant. But note, that not every meiosis involves a crossing over between the genes observed!

The frequencies of recombinants arising from crossing over are always less than 50% (Figure 81.).

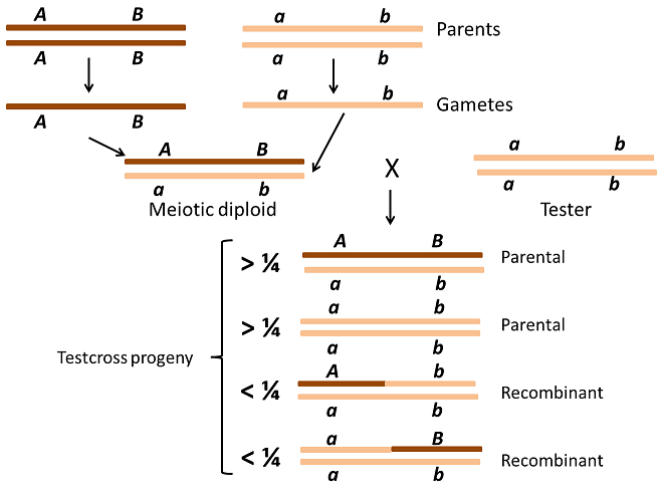


Figure 81. Recombination between linked genes

4.3 Transformation

Molecular biologists and geneticists call transformation the process, when ***an organism takes up DNA from the surrounding and integrates it into the genome, thereby changing its own genotype*** permanently. The transforming DNA can be of the same species or of another species, as well.

Transformation was first discovered in the bacterium *Streptococcus pneumoniae* in 1928; since then, the technique has been developed for several microbes and model organisms (Figure 82.). "Transformation" may also be used to describe the insertion of new genetic material into plant and animal cells. However, to avoid misunderstanding, in animals the process is usually called transfection. (Transformation of animal cells means a progression to a cancerous state.) Several methods are known to transfer DNA into plant cells. They will be described in Section 9.3.

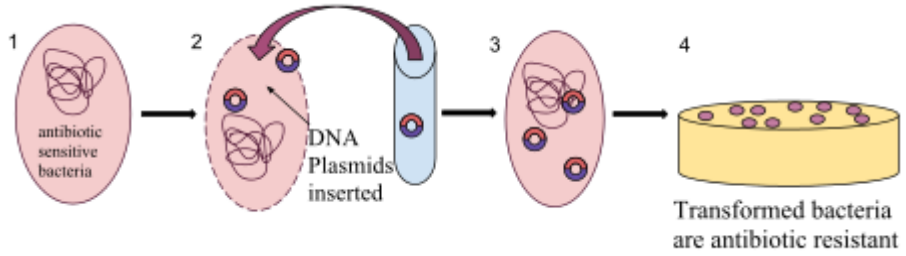


Figure 82. Bacterial transformation

Before transformation, bacterial cells are sensitive to antibiotics. Chemical or physical treatment of the cells make them competent to take up DNA. The plasmid, this small circular DNA, carries an antibiotic resistance gene. Transformed cells that have taken up the plasmid became resistant, and can grow on selective medium containing antibiotics.
https://commons.wikimedia.org/wiki/File:Artificial_Bacterial_Transformation.svg
Amunroe13, CC BY-SA 3.0 <https://creativecommons.org/licenses/by-sa/3.0>, via Wikimedia Commons

4.4 Large-scale chromosomal changes

Changes in chromosome structure

Chromosome mutations are large-scale chromosomal changes that can be divided into two groups: ***changes in chromosome structure*** and ***changes in chromosome number***. Chromosome mutations can be detected by microscopic examination and genetic analysis. They have been best characterized in eukaryotes.

Chromosome mutations can be of interest from different biological aspects: They can show how genes act in concert on a genomic scale. They reveal important features of meiosis and chromosome architecture, and can be useful tools for the experimental manipulation of the genome. Chromosome mutations can be a source of insight into evolutionary processes as well.

Changes in chromosome structure, also called ***rearrangements***, involve four types of events (Figure 83.): ***Deletion*** is the name of event when a chromosome segment is lost. Doubling a portion of a chromosome creates ***duplication***. When the orientation of a segment is reversed within the chromosome, the event is called ***inversion***. Movement of a chromosome segment to a different chromosome is called ***translocation***.

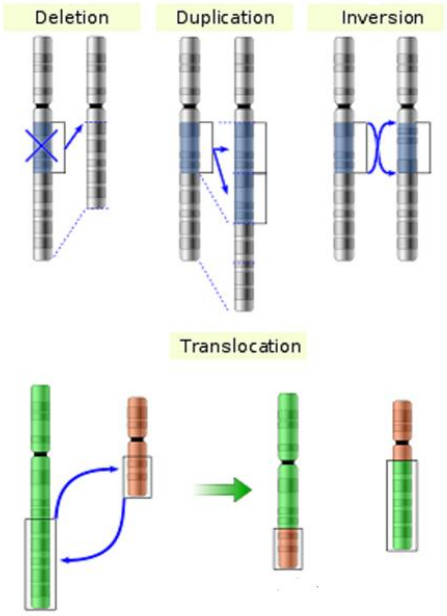


Figure 83. Chromosome rearrangements

Based on: https://commons.wikimedia.org/wiki/File:Chromosomes_mutations-en.svg
GYassineMrabetTalk ✉ This W3C-unspecified vector image was created with Inkscape., Public domain, via Wikimedia Commons

Only those rearrangements produce viable products, which retain one centromere and two telomeres of a chromosome. Chromosomes lacking a centromere (***acentric fragment***) will be lost during cell division, because they cannot attach to the spindle. If the rearrangement produces a ***dicentric chromosome***, it will be pulled to opposite poles at anaphase, and break. Chromosome fragments will not be incorporated into progeny cells in either case. Chromosomes ***lacking normal telomeres*** cannot be replicated properly.
Balanced chromosomal rearrangements change the order of genes on the chromosomes, but do not change the amount of DNA. These are the inversion and the reciprocal translocation.
Unbalanced rearrangements change the gene dosage of a chromosome segment. This is true for the deletion and the duplication.
Chromosomal rearrangements are often produced by breakage and rejoining of chromosomes. Crossing over between repetitive sequences can also produce each of the four types of rearrangements.

Inversions

An inversion is said to be ***paracentric*** if the centromere is outside the inverted chromosome segment. Inversions spanning the centromere are called ***pericentric***. Inversions are balanced rearrangements, thus individuals with inversions usually show normal phenotype. An exception is when the production of the new inversion involved breakage within a gene.

Inversion heterozygotes (diploid cells carrying a normal and an inversion chromosome) can be identified cytologically. During meiosis, paring of the homologous chromosomes requires that one of them twists once at the ends of the inversion. This is visible as an ***inversion loop*** in the microscope (Figure 84.).

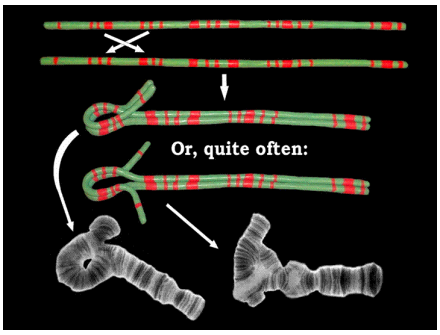


Figure 84. Heterozygous inversion loops in polytene chromosomes

https://commons.wikimedia.org/wiki/File:Inversioncartoon.gif
The original uploader was Axarus at English Wikipedia., CC BY-SA 1.0 <https://creativecommons.org/licenses/by-sa/1.0>, via Wikimedia Commons

Crossing over within the inversion loop has an effect on the meiotic products (Figure 85.). In the case of a ***paracentric inversion***, a dicentric bridge and an acentric fragment are produced. Tension breaks the dicentric fragment at anaphase I forming two chromosomes with terminal deletions. The acentric fragment is lost. Gametes that contain the deletion products or zygotes that form from these gametes will probably be inviable. The overall result is a drastically lower frequency of viable recombinants.
In the case of a ***pericentric inversion***, the chromosomes that have engaged in crossing-over separate normally, but a portion of the meiotic products carries duplication-deletion. If a gamete carrying a crossover chromosome is fertilized, the zygote dies because of gene inbalance.

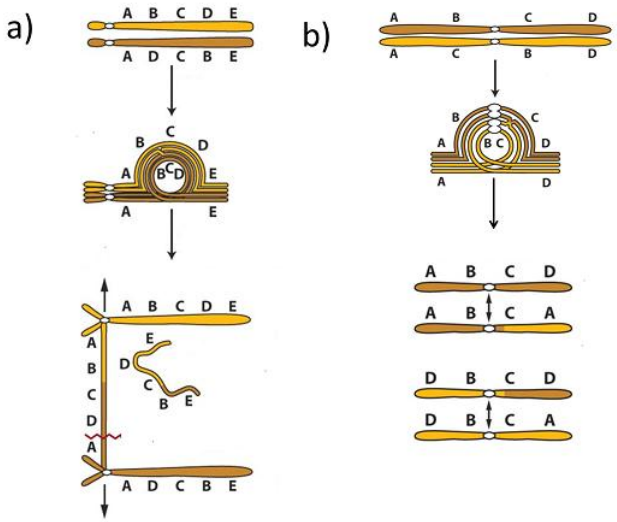


Figure 85. The effect of crossing over in inversion heterozygotes

a) In a paracentric inversion heterozygote, crossing over results in a dicentric chromosome and an acentric fragment. b) In pericentric inversion heterozygotes, only 50% of the meiotic products carry a complete gene set.

In both cases, only meiotic products that have not engaged in crossing over can produce viable progeny. Thus, ***crossing over in inversion heterozygotes leads to reduced recombinant frequencies and reduced fertility.***

Reciprocal translocations

Reciprocal translocations between nonhomologous chromosomes are balanced rearrangements, thus they usually do not cause phenotypic changes, unless the breakpoint for the translocation occurred within a gene.

Reciprocal translocations can ***change the arm ratios*** of chromosomes significantly (e.g. see Figure 83. on page 61), and create apparent linkage of genes whose normal loci are on separate chromosomes (***pseudolinkage***).

Translocations are ***important factors in the evolution of genomes***.

Meiosis in translocation heterozygotes produces characteristic configurations. ***The paring configuration is cross-shaped*** (Figure 86.).

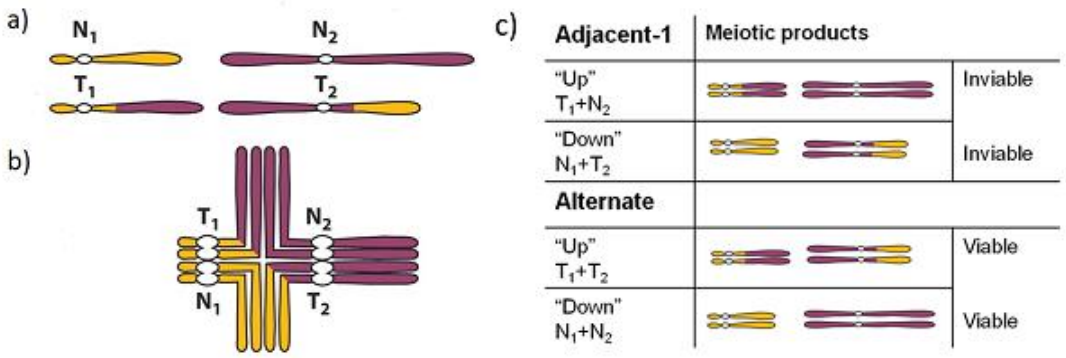


Figure 86. Meiosis in translocation heterozygotes

a) The normal (N) and translocated (T) chromosomes to be paired; b) The charactristic pairing configuration; c) The possible results of the meiotic segregation

There are two common patterns of segregation: In adjacent-1 segregation each of the two meiotic products is deficient for a chromosome arm, and has a duplication of another arm. These products are inviable. In alternate segregation, a completely normal product and a product with the translocated chromosomes are produced. They are both viable. Thus, ***the overall result of the meiosis of a translocation heterozygote is semisterility***.

Semisterility is typical for translocation heterozygotes, but defined differently for plants and animals. In plants, meiotic products that contain imbalanced chromosomes generally abort at the gametic stage. In animals, gametes with imbalanced chromosomes are capable of fertilization, but the zygote dies.

Deletions

A deletion is simply the loss of a part of a chromosome. A single break results in a ***terminal deletion***. Breaking chromosomes twice can produce an ***interstitial deletion*** (Figure 87.).

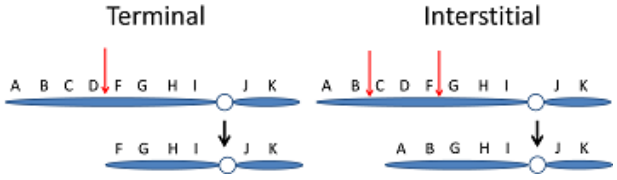


Figure 87. Types of deletion

The effect of deletions on viability depends on their size.

1. A small deletion within a gene (also called ***intragenic deletion***) can be viable, if the homozygous null mutation of that gene is viable. Intragenic deletions can be distinguished from point mutations, because genes with such deletions ***never revert*** to wild-type.
2. Loss of several or many genes (***multigenic deletion***) have more severe consequences. If they are made homozygous by inbreeding, they are always lethal. Even individuals heterozygous for a multigenic deletion may not survive. This can be explained by a ***disruption of normal gene balance***, or expression of deleterious recessive alleles (***pseudodominance***) (Figure 88.).

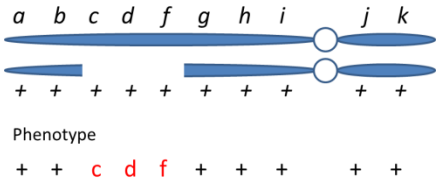


Figure 88. Pseudodominance in a deletion heterozygote

Crossing over of chromosomes cannot happen in the area of a deletion, therefore, heterozygous deletions act as ***inhibitors of recombination*** in the affected area.

Deletion heterozygotes carrying a small deletion can be identified under the microscope. At the area of deletion, the normal homolog has no partner for meiotic pairing, and thus protrudes as a visible ***deletion loop***.

Duplications

The process of chromosome mutation sometimes creates an extra copy of some chromosome regions. The duplicate regions can be located adjacent to each other in the same orientation (called ***tandem duplication***), or in opposite orientation (***reverse duplication***). If the extra copy is located elsewhere in the genome, it is called an ***insertional duplication***.

Meiotic pairing of tandem and reverse duplications can ***result in interesting configurations*** (Figure 89a.).

Inbreeding of individuals carrying tandem duplication may result in a duplication homozygote in which there are already 4 copies of the duplicated region. In such individuals, asymmetric meiotic pairing and recombination may lead to ***triplication*** (Figure 89b.).

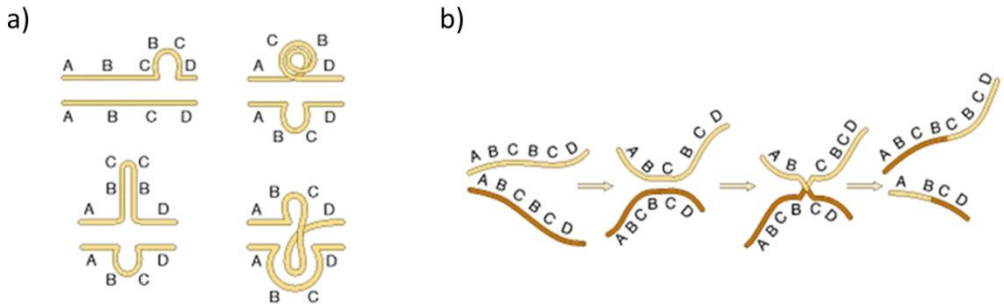


Figure 89. Meiotic problems caused by duplication

a) Pairing configurations; b) “Slippage” during pairing in a duplication homozygote and the subsequent recombination resulted a triplication and a normal chromosome.

When a gene has multiple copies in the genome, the extra copies are free to mutate, because a single copy of the gene can provide its basic function. Therefore, duplications provide raw material for genome changes during the evolution.
When individuals carrying a duplication show some phenotype, it is usually due to genome imbalances.

Changes in chromosome number

The number of chromosomes varies greatly in flora and fauna. Changes in chromosome number are of special importance for plant breeders who have routinely manipulated chromosome number to improve commercially important agricultural crops. This is an important area in human health care as well, because a group of common genetic disorders results from the abnormal number of chromosomes.

Changes of chromosome number are of two basic types: Changes in whole chromosome sets lead to a condition called ***aberrant euploidy***. Changes in parts of chromosome sets result in a condition called ***aneuploidy***.
Organisms with multiples of the basic chromosome set (genome) are referred to as euploid. Higher organisms (animals and plants) ***are usually diploid*** (they have two chromosome sets: ***2n***), and ***form haploid gametes*** (which carry one chromosome set: ***n***). Other organisms (e.g.: fungi, algae) are essentially haploid.
Aberrant euploidy means that an organism has more (***polyploid***) or fewer (***monoploid***) chromosome sets than the normal number.

Monoploids

Male bees, wasps, and ants are monoploid (Figure 90.). They develop by ***parthenogenesis***: the egg develops into an embryo without fertilization. (In these groups, male individuals produce gametes by mitosis.)

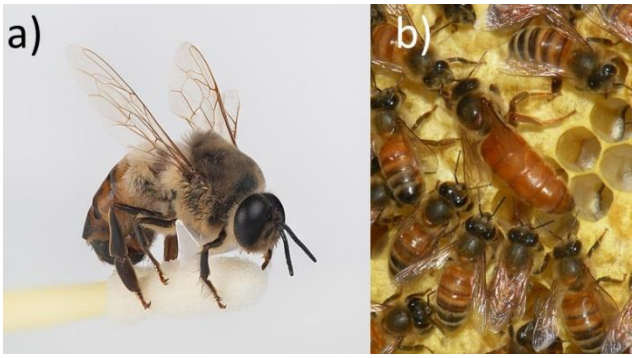


Figure 90. Bee family members

a) A drone honey bee (*Apis mellifera*). Drones are monoploids, and descend only from their mother, the queen bee. b) Queen bee and workers. The queen be is the only fertile female in the beehive. It is larger than the workers, and has fully developed reproductive system.
https://commons.wikimedia.org/wiki/File:Drone_bee_(32-image_macro_stack).jpg
Guillaume Pelletier, CC BY-SA 4.0 <https://creativecommons.org/licenses/by-sa/4.0>, via Wikimedia Commons
https://commons.wikimedia.org/wiki/File:Apis_mellifera_(queen_and_workers).jpg
Jessica Lawrence, Eurofins Agroscience Services, Bugwood.org, CC BY 3.0 <https://creativecommons.org/licenses/by/3.0>, via Wikimedia Commons

In most species, monoploid zygotes (unfertilized eggs) fail to develop. The reason is that deleterious recessive mutations, which are "masked" in a diploid organism, are automatically expressed in a monoploid individual. Due to this "genetic load", monoploids that develop to

adulthood are abnormal. If they reach reproductive age, they prove to be steril, because in lack of pairing partners, their chromosomes cannot segregate properly in meiosis.
Monoploid plants are routinely generated by tissue culture. We will return to this topic in the "Agricultural Biotechnology" chapter.

Polyploids

Polyploidy is very common in plants but rarer in animals.

Polyploid animals: Examples of polyploidy are more common in non-vertebrate animals, such as ***flatworms***, ***leeches***, and ***brine shrimps***. In these groups, reproduction occurs by parthenogenesis. Within vertebrates, polyploid ***fishes and amphibians*** occur commonly. For example, salmonids (*Salmonidae*) and many cyprinids (*Cyprinidae*) show stable polyploidy. The ***Salmonidae family*** of fishes is a familiar example of a group that appears to have originated through polyploidy. The biomedically important genus *Xenopus* contains several polyploid species. Polyploid ***reptiles***, like e.g. lizards, are also quite common. Polyploid frogs and toads reproduce sexually, whereas polyploid salamanders and lizards are parthenogenetic.

The sterility of triploids has been commercially exploited both in animals and plants. Oysters are unpalatable during the spawning season. To prevent spawning, sterile oysters are now cultivated by crossbreeding tetrapolid and diploid oysters. The resulting ***triploid oysters*** do not spawn, and are palatable the whole year round.

Polyploid plants: In the flora, an increase in the number of chromosome sets has been an ***important factor in the origin of new species***. Upon examination of the frequency distribution of haploid chromosome number in dicotyledonous plants, a striking excess of even-numbered values can be seen in the higher ranges. This suggests ancestral polyploidization, because doubling and redoubling of a number can give rise only to even numbers.

Polyploid plants, as a whole and in their component parts, are often larger than their diploid relatives (Figure 91.).

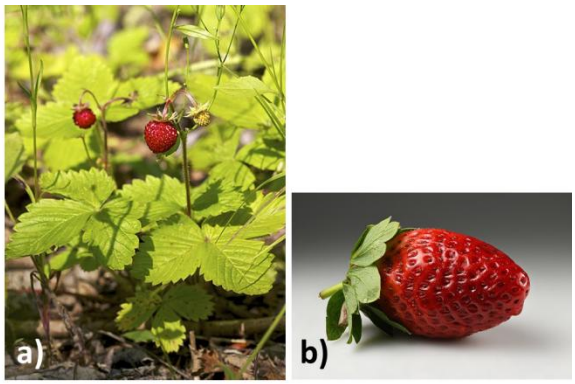


Figure 91. Strawberry fruits

a) The species *Fragaria vesca*, commonly known as wild strawberry, was the first strawberry ever cultivated. This species is diploid. b) The garden strawberry, a worldwide cultivated hybrid species (*Fragaria x ananassa*), is octoploid.

https://commons.wikimedia.org/wiki/File:Fragaria_vesca_LC0389.jpg
Jörg Hempel, CC BY-SA 3.0 DE <https://creativecommons.org/licenses/by-sa/3.0/de/deed.en>, via Wikimedia Commons
https://commons.wikimedia.org/wiki/File:Garden_strawberry_(Fragaria_%C3%97_ananassa)_single2.jpg
Ivar Leidus, CC BY-SA 4.0 <https://creativecommons.org/licenses/by-sa/4.0>, via Wikimedia Commons

Polyploids can be classified as autopolyploids and allopolyploids. In case of ***autopolyploids***, the multiple sets of chromosomes originate from one species. The related chromosomes are said to be ***homologous***. The multiple chromosome sets of ***allopolyploids*** are derived from two or more species. Here, the related chromosomes correspond only partially to each other and are said to be ***homeologous***.

Autopolyploid plants

Autotriploids: Autotriploids arise spontaneously in nature, but also geneticists can construct them by ***crossing tetraploid (4n) and diploid (2n) plants***. ***Triploids are characteristically sterile***, due to the problem of pairing of three homologous chromosomes in meiosis (Figure 92.).

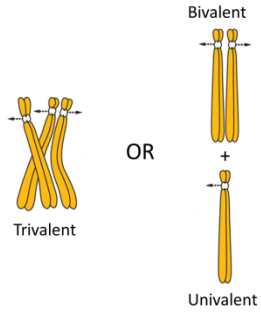


Figure 92. Pairing options of homologous chromosomes in triploids

Whichever pairing mechanism is realized, the likelihood that gametes with complete chromosome sets arise (diploid or haploid) is very low. Aneuploid gametes generally do not give rise to viable offspring.
The sterility of triploids can be an advantage: commercially available bananas are seedless autotriploids (3n=33), and there are seedless watermelons as well (Figure 93.).

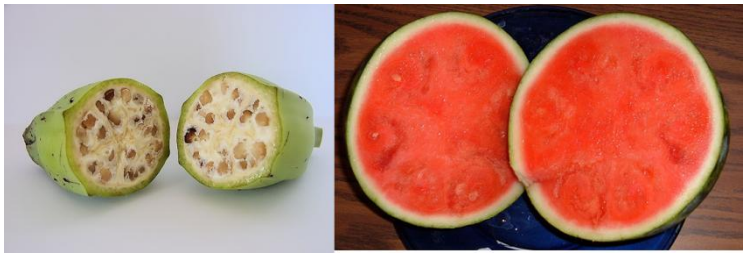


Figure 93. The advantage of being triploid

The fruit of a diploid banana, which has numerous large and hard seeds, could hardly be enjoyable (left). Seedless, triploid watermelon (right).
https://commons.wikimedia.org/wiki/File:Inside_a_wild-type_banana.jpg
Warut Roonguthai, CC BY-SA 3.0 <https://creativecommons.org/licenses/by-sa/3.0>, via Wikimedia Commons
https://commons.wikimedia.org/wiki/File:Watermelon_seedless.jpg
Scott Ehardt, Public domain, via Wikimedia Commons

Autotetraploids: Autotetraploids arise by ***doubling the chromosome sets of diploid organisms***. This can occur spontaneously, but can also be induced artificially. Treating diploid somatic tissues of plants with ***colchicine***, a commonly used antitubulin agent, for one cell cycle, leads to tetraploids.

The fertility of autotetraploids depends on how their chromosomes pair in meiosis (Figure 94.).

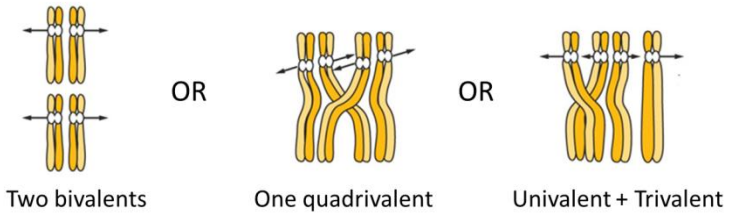


Figure 94. Pairing options of homologous chromosomes in tetraploids

If the preferred pairing mode is formation of two bivalents or one quadrivalent, functional diploid gametes can be produced. The fusion of gametes at fertilization regenerates the tetraploid state.
A number of autotetraploid plants have been produced by breeding. The attraction of tetraploid varieties is the ***larger flower size, bigger fruit***, or the ***higher crop yield***.

Allopolyploid plants

Allopolyploid plants can be created by crossing related species and doubling the chromosomes of the hybrid or by fusing diploid cells.

The prototypic allopolyploid was an ***allotetraploid*** synthesized by G. Karpachenko in 1928 (Figure 95.).

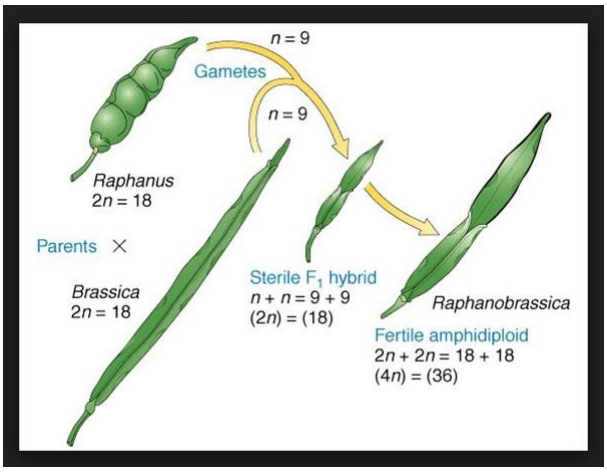


Figure 95. The origin of *Raphanobrassica*

https://commons.wikimedia.org/wiki/File:Brassicoraphanus.jpg
Raphanobrassica.jpg, CC BY-SA 4.0 <https://creativecommons.org/licenses/by-sa/4.0>, via Wikimedia Commons

His aim was to make a fertile hybrid that would have cabbage leaves (*Brassica*) and radish roots (*Raphanus*). The species are related closely enough to allow intercrossing, but the resulted (n_1+n_2) hybrid was sterile. An accidental chromosome doubling occurred in the hybrid plant, presumably in a tissue that eventually became a flower and produced gametes. This led to the production of a fertile ($2n_1$+$2n_2$) allopolyploid, the ***Raphanobrassica***. This kind of allopolyploid is also called ***amphidiploid***, or doubled diploid.

In nature, allopolyploidy have been a major force in evolution of new plant species. One example is shown by the genus Brassica (Figure 96.).

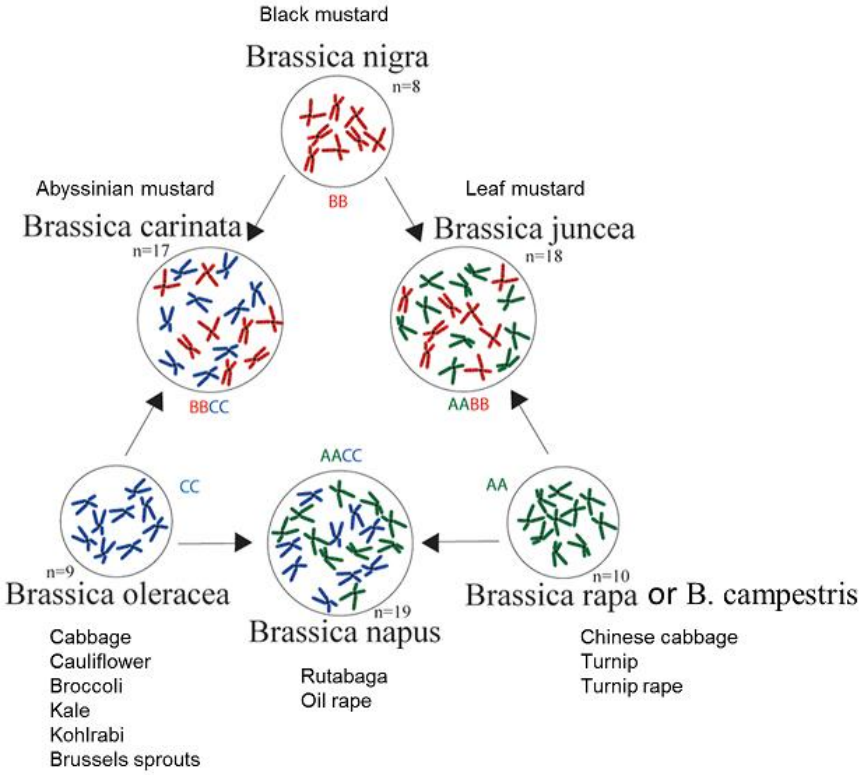


Figure 96. The triangle of U

The diagram presents a theory about the evolution and relationship among the most commonly known six members of the genus Brassica. This theory was first published in 1935, by Woo Jang-choon, a Korean-Japanese botanist, who published under the Japanese name Nagaharu U. Cultivars of the species are above/below the Latin names.

Based on: https://commons.wikimedia.org/wiki/File:Triangle_of_U_Simple1.PNG

A particularly important natural allopolyploid is the ***bread wheat, Triticum aestivum***. Bread wheat is an ***allohexaploid*** (6n=42), whose genome is ***composed of three ancestral genomes*** (Figure 97.). At meiosis, special mechanism ensures that ***chromosome pairing always occurs between homologous chromosomes*** of the same ancestral genome. Hence, in bread wheat meiosis, there are always 21 bivalents. The plant is completely fertile.

Allopolyploidy has been important in the production of modern crop plants. ***Triticale*** is a synthetic ***amphidiploid between wheat*** (*Triticum*, 6n=42) ***and rye*** (*Secale*, 2n=14). Hence, for *Triticale*, 2n = 2 x (21+7) = 56. *Triticale* combines the high yield of wheat with the ruggedness of rye.

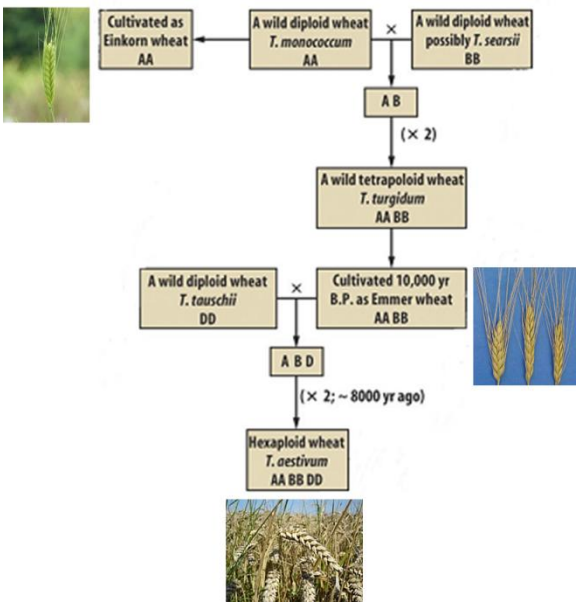


Figure 97. The proposed evolutionary history of modern wheat

The diploid Einkorn wheat and the tetraploid, domesticated Emmer wheat are still cultivated beside the hexaploid modern wheat.
https://commons.wikimedia.org/wiki/File:Triticum_monococcum0.jpg; https://creativecommons.org/licenses/by-sa/3.0/deed.en
https://commons.wikimedia.org/wiki/File:Usdaemmer1.jpg
https://commons.wikimedia.org/wiki/File:Wheat_close-up.JPG
User:Bluemoose, CC BY-SA 3.0 <http://creativecommons.org/licenses/by-sa/3.0/>, via Wikimedia Commons

Figure 98. Wheat, rye, and Triticale

https://commons.wikimedia.org/wiki/File:Wheat,_rye,_triticale_montage.jpg

Animal hybrids

The most remarkable hybrid is the ***mule*** (Figure 99.): the ***offspring of a female horse*** (*Equs caballus*; 2n = 64) ***and a male donkey*** (*Equs asinus*; 2n = 62). Mules are ***sterile***.

The liger is a hybrid offspring of a male lion (*Pathera leo*) and a female tiger (*Panthera tigris*). Ligers are the biggest felines.

Figure 99. Animal hybrids: a mule and ligers

https://commons.wikimedia.org/wiki/File:Juancito.jpg
w:User:Dario u / User:Dario urruty, Public domain, via Wikimedia Commons
https://commons.wikimedia.org/wiki/File:Liger_couple.jpg
Hkandy, CC BY-SA 3.0 <https://creativecommons.org/licenses/by-sa/3.0>, via Wikimedia Commons

Aneuploids

An aneuploid is an individual whose chromosome number ***differs from the wild-type by part of a chromosome set***. Typically, only one or a few chromosomes are missing or are in excess.

Nomenclature for aneuploid conditions: In diploids: 2n + 1 = ***trisomic***; 2n – 1 = ***monosomic***; 2n – 2 = ***nullisomic***. In haploids: n + 1 = ***disomic***.

Plants tolerate aneuploidy much better than animals. Among the aneuploids, trisomics occur the most commonly. Trisomy was first observed in jimsonweed (*Datura stramonium*) by Alfred Blakeslee (Figure 100.). The plant has 12 chromosomes. Blakeslee obtained 12 different strains, each of which had the normal 12 chromosome pairs plus an extra representative of one pair. Each strain was phenotypically distinct from the others. This suggested that each chromosome in the genome carries qualitatively different genes.

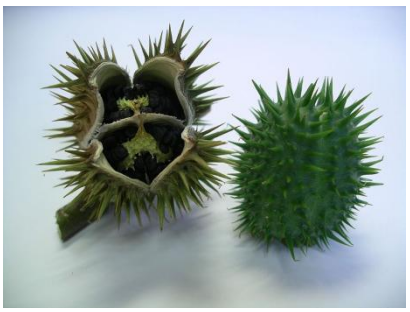


Figure 100. Jimsonweed seedpods

https://commons.wikimedia.org/wiki/File:Thorn-apple.jpg
Corin Royal Drummond from Ajijic, Jalisco, Mexico, CC BY-SA 2.0 <https://creativecommons.org/licenses/by-sa/2.0>, via Wikimedia Commons

5. Extranuclear Inheritance

5.1 The structure and origin of chloroplasts and mitochondria

The structure of mitochondria

Mitochondria (Figure 101.) are present in every eukaryotic cell. They are described as the ***power plants of cells***; the amount of energy produced by mitochondria greatly surpasses that of produced by glycolysis, thus the appearance of mitochondria has contributed to the development of complex eukaryotic organisms fundamentally.

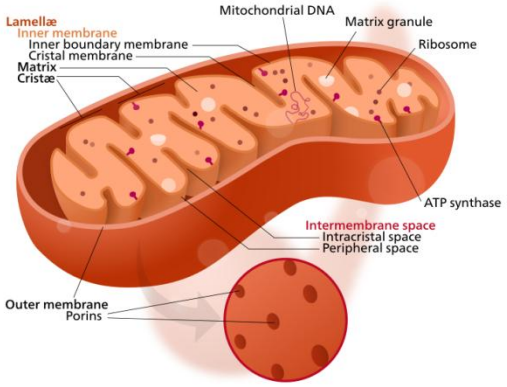


Figure 101. The structure of mitochondria

https://commons.wikimedia.org/wiki/File:Mitochondrion_structure.svg
Kelvinsong; modified by Sowlos, CC BY-SA 3.0 <https://creativecommons.org/licenses/by-sa/3.0>, via Wikimedia Commons

Mitochondria are bounded by two highly specialized membranes that have very different functions. The ***outer membrane*** is characterized by average composition and permeability. Molecules passing through the outer membrane enter the ***intermembrane space***. The ***inner membrane*** is highly impermeable to ions, and forms a series of infoldings called ***cristae***. The inner membrane contains the protein complexes of the ***respiratory chain***, as well as many ***transport proteins***. The space surrounded by the inner membrane is called ***matrix***. It contains ***mitochondrial DNA (mtDNA)***, ribosomes, tRNA, and enzymes for DNA duplication, transcription, and translation. In addition, the matrix contains the enzymes of the ***citric acid cycle and ß-oxidation***.

The structure of chloroplasts

Chloroplasts (Figure 102.) are the most typical organelles of plants. They are surrounded by a highly permeable ***outer membrane***, and have a much less permeable ***inner membrane***, in which membrane transport proteins are embedded. There is a narrow ***intermembrane space*** between them. Together, these membranes form the chloroplast envelop.

Inside the inner membrane, there is a large space called ***stroma***. The stroma contains many metabolic enzymes, ribosomes, RNAs, and the ***chloroplast DNA (cpDNA)***.

A third distinct membrane, the ***thylakoid membrane*** forms a set of flattened, disc-like sacks called thylakoids within the stroma. The thylakoids tend to stack to form ***grana***. The lumens of the thylakoids are connected with each other, and so define a third internal compartment, the ***thylakoid space***.

The job of chloroplasts is to capture energy from the sunlight, and to use it to fix carbon. The ***electron transport chain*** and the ***ATP synthesizing enzymes*** of the ***"light reactions"*** reside in the thylakoid membrane, while ***carbon-fixation*** in the ***"dark reactions"*** continues in the stroma.

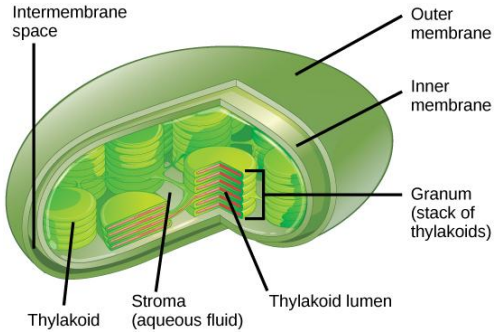


Figure 102. The structure of chloroplasts

https://commons.wikimedia.org/wiki/File:Figure_08_01_05.png
CNX OpenStax, CC BY 4.0 <https://creativecommons.org/licenses/by/4.0>, via Wikimedia Commons

The endosymbiosis theory

The theory of ***endosymbiosis*** suggests that ***certain eukaryotic organelles evolved from prokaryotic organisms***, which entered into symbiosis with an ancestor of eukaryotic cells, "the protoeukaryote" (Figure 103.). Aerobic bacteria established a stable residence inside the cytoplasm of a primitive eukaryotic-like cell, providing the cell with energy in return for a protected environment and an easily obtainable source of nutrients. This symbiotic relationship created the forerunner of the ***mitochondrion*** in the modern eukaryotic cell. Similarly, a primitive eukaryote gained photosynthetic properties after the endosymbiotic uptake of an oxygen producing phototroph, the forerunner of the modern ***chloroplast***.

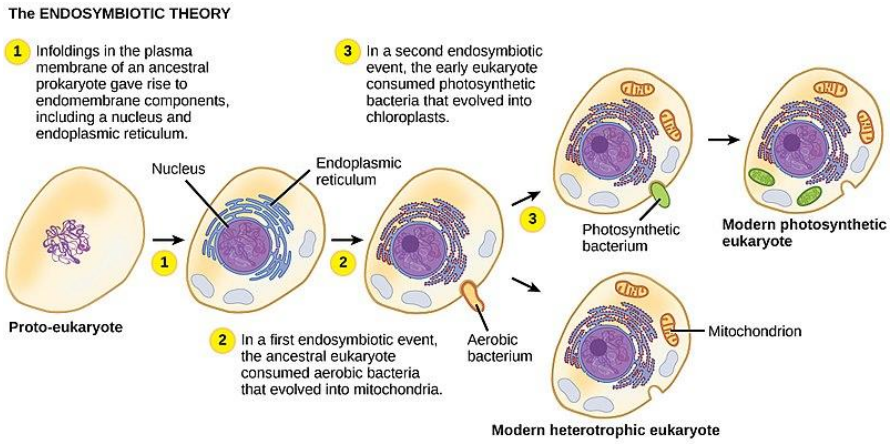


Figure 103. Simple demonstration of the endosymbiosis theory

Eukaryotic cells with mitochondria arose ~2 billion years ago. Chloroplasts are derived from the fusion of cyanobacteria with the common ancestor of algae and higher plants.
https://commons.wikimedia.org/wiki/File:Endosymbiotic_theory.jpg
Phil Schatz, CC BY-SA 4.0 <https://creativecommons.org/licenses/by-sa/4.0>, via Wikimedia Commons

Mitochondria probably arose from a major group of *Bacteria* called ***Proteobacteria***.
The primary endosymbiosis that has led to the development of chloroplasts occurred between ***cyanobacteria*** and primitive algae already possessing mitochondria. Such symbiosis has occurred several times during the evolution of plants. Red algae, glaucophytes, and green algae originated by primary endosymbiosis. Green algae became the ancestors of terrestrial plants. Several taxa acquired their chloroplasts by establishing symbiosis with an alga already having a chloroplast.

The endosymbiosis theory has several ***"evidences"***, such as:
Mitochondria and chloroplasts → have their own ***genome***;
→ have ***prokaryotic ribosomes***;
→ are capable of ***protein synthesis***;
→ are ***surrounded by a double-membrane***;
→ and are not involved in the vesicular transport that connects other organelles.

During the evolution, both organelles lost their original independence. Most of the genes encoding the components (proteins, lipids) needed to form the organelles have been transferred to the host genome. For example, the human mitochondrial genome encodes 13 mitochondrial proteins, whereas over 600 other mitochondrial proteins are encoded in the nucleus, and imported from the cytoplasm after being synthesized on cytoplasmic ribosomes.

5.2 The extranuclear genome

Copy number and organization of DNA

Organelle genomes are neither haploid nor diploid. Eukaryotic cells contain several, or even many mitochondria, and each mitochondrion contains several copies of the mitochondrial DNA. The number of organelles, as well as the copy number of DNA within the individual organelles, can vary from cell to cell. The same is true for chloroplasts; ***mitochondrial and chloroplast DNA are present in multiple copies*** in these organelles (Table IV.).

Organism	Tissue / cell type	DNA molecules / organelle	Organelles/ cell	Organelle DNA as % of total cellular DNA
MITOCHONDRIAL DNA				
Rat	liver	5-10	1000	1
Yeast	vegetative	2-50	1-50	15
Frog	egg	5-10	10^7	99
CHLOROPLAST DNA				
Chlamydomonas	vegetative	80	1	7
Maize	leaves	20-40	20-40	15

Table IV. Mitochondrial and chloroplast DNA content of different cell types

(Content based on: Alberts, Johnson, Lewis, Morgan, Raff, Roberts, Walter: Molecular Biology of the Cell, page 802.)

Mitochondrial and chloroplast DNAs contained within the matrix or stroma of these organelles are usually distributed in several clusters, called ***nucleoids***, and are ***not associated with histone proteins***. In both of these properties, they resemble the hereditary material of bacteria.

Mitochondrial DNA molecules range in size from less than 6kb (*Plasmodium falciparum*) to more than 300kb (plants). Most mtDNAs are ***circular*** molecules.
The chloroplast genome of plants ranges in size from 120 to 200kb, and is ***circular*** in all organisms examined thus far.

Overview of the mitochondrial genome

Although the sizes of mitochondrial genomes vary largely, the genome sizes do not correlate well with the number of proteins encoded in them. The smaller mitochrondrial genomes tend to be more densely packed; in the larger mitochondrial genomes there is more noncoding DNA.
Mitochondrial genomes seem to serve two main functions. They code for some proteins working in the ***electron transport chain***, and code for some proteins, all the tRNAs, and both rRNAs necessary for ***mitochondrial protein synthesis***. However, the remaining necessary components for these functions are encoded by nuclear genes indicating a ***division of labor*** between the nuclear and mitochondrial DNA. Interestingly, some specific subunits are encoded by mtDNA in one organism but by nuclear DNA in another. This shows an ***evolutionary transposition of information*** between these organelles.

Due to its relatively small size, the human mitochondrial genome was sequenced early (Figure 104.). Analysis of the sequence has revealed some surprising features.

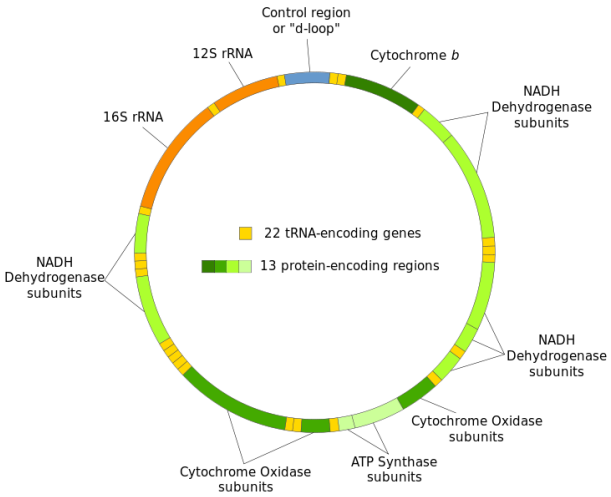


Figure 104. Map of the human mitochondrial genome

https://commons.wikimedia.org/wiki/File:Mitochondrial_DNA_en.svg
derivative work: Shanel (talk)Mitochondrial DNA de.svg: translation by Knopfkind; layout by jhc, CC BY-SA 3.0 <http://creativecommons.org/licenses/by-sa/3.0/>, via Wikimedia Commons

1. The 16,569 nucleotides long genome is ***densely packed***; it contains 2rRNA genes, 22tRNA genes, and 13 protein-coding sequences.
2. The normal ***codon-anticodon pairing rules are relaxed***; many tRNAs recognize all four nucleotides in the third position. This allows protein synthesis with fewer tRNAs in the mitochondrion than in the cytoplasm.
3. There are some ***differences between the nuclear and the mitochondrial codon usage***.

The relaxed codon usage (“extra wobble”) and the slight deviations from the “universal” genetic code proved to be general features of the mitochondrial coding. A further interesting and surprising feature of the mitochondrial genome is that some plant and fungal (including yeast) mitochondrial genes contain ***introns***.

An overview of the chloroplast genome

The organization of chloroplast genome is very similar in plants, although the size varies from species to species (Figure 105.).

Chloroplast genes are involved mainly in ***protein synthesis and photosynthesis***. Typically, the known proteins encoded in the chloroplast are part of larger protein complexes that also contain one or more subunits encoded in the nucleus. A good example for this is the ribulose-1-5-biphosphate carboxylase/oxygenase (***rubisco***), an enzyme involved in the first major step of carbon fixation. It is probably the most abundant enzyme on Earth. Proteins of the large subunit of this enzyme are encoded by the chloroplast DNA, while those of the small subunit are encoded by nuclear genes.

A further characteristic of the chloroplast genomes is the presence of ***inverted repeats***. These were found in the cpDNA of virtually all species of plants.

Similarly to the mtDNA, genes of the cpDNA may contain ***introns***.

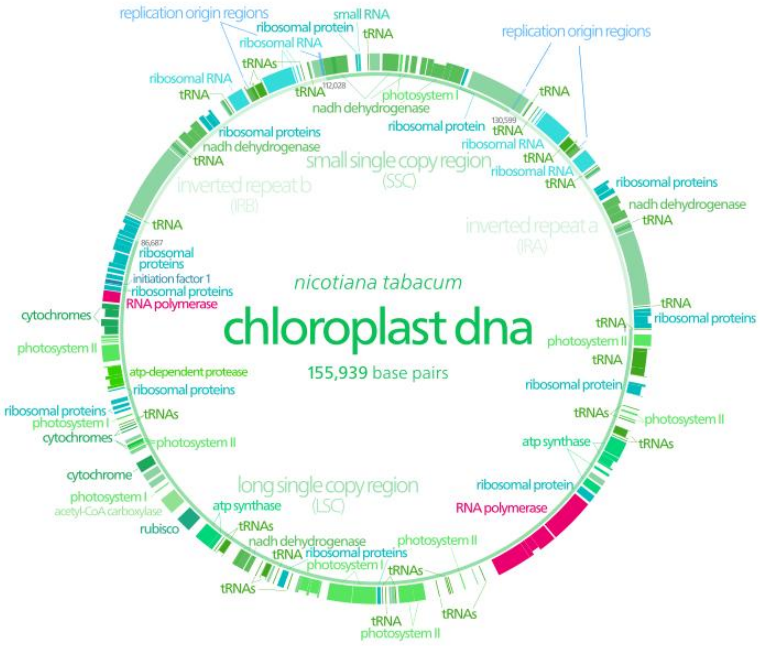


Figure 105. Gene map of tobacco chloroplast DNA

https://commons.wikimedia.org/wiki/File:CtDNA.svg
Kelvinsong, CC BY-SA 3.0 <https://creativecommons.org/licenses/by-sa/3.0>, via Wikimedia Commons

There are striking similarities between the genomes of chloroplasts and bacteria.

Comparison of genome sequences allowed drawing some important conclusions:

1. Sequence comparisons support the hypothesis that ***higher plants arose from photosynthetic bacteria***.
2. ***The chloroplast genome changes slowly***; it has been stably maintained for several hundred million years.
3. Many of the ***original bacterial genes are now present in the nuclear genome*** of higher plants.

5.3 Examples of extranuclear inheritance

Leaf variegation in Mirabilis jalapa

In 1909, Carl Correns published some surprising results from his studies on four-o'clock plants (*Mirabilis jalapa*). This plant has variegated, green and white branches, and flowers may form on any branch. (For variegated phenotype, see Figure 59b. on page 49.). Correns intercrossed a variety of different combinations by transferring pollen from one flower to another. His results are summerized in Table V.

Phenotype of branch bearing egg parent	Phenotype of branch bearing pollen parent	Phenotype of progeny
White	White	White
White	Green	White
White	Variegated	White
Green	White	Green
Green	Green	Green
Green	Variegated	Green
Variegated	White	Variegated, green or white
Variegated	Green	Variegated, green or white
Variegated	Variegált	Variegated, green or white

Table V. Results from intercrossing different *Mirabilis jalapa* plants

Surprisingly, the phenotype (actually, genotype) of the maternal parent is solely responsible for determining the phenotype of all progeny. The phenotype (genotype) of the male parent seems to be completely irrelevant. For this reason, this pattern of inheritance is called ***maternal inheritance***.

The green color of leaves is caused by the chloroplast dye chlorophyll. Some of the chloroplast genes control the production and localization of chlorophyll. Mutations in these genes lead to the formation of white leaf and shoot segments.

Chloroplasts are localized in the cytoplasm of cells, and the zygote receives the bulk of the cytoplasm from the maternal parent. The phenotype of the offspring is determined by the genotype of the cpDNA, trasmitted by the cytoplasm of the egg.

Variegated branches apparently produce three types of eggs: some with white chloroplasts (→ white shoot in progeny), some with green chloroplasts (→ green shoot in progeny), and some with both kinds of chloroplasts. The egg type containing both green and white chloroplasts transmits both chloroplast types to the zygotes. In subsequent mitotic divisions, some cellular process causes similar chloroplasts segregate together in some cell lines. This results in variegated phenotype in the offspring.

Cells with mixtures of organelle genotypes are called ***cytohets*** ("cytoplasmic heterozygote"). A cell that is cytohet may have, say 60% organelles carrying "*A*" allele (wild-type) and 40% organelles with "*a*" allele (mutant). When such cells divide, organelles of different genotypes may be separated in the offspring randomly. This partitioning usually requires several cell divisions. This type of genetic segregation is called ***cytoplasmic segregation*** (Figure 106.). Cytoplasmic segregation occurs in dividing asexual cells, but it is unrelated to mitosis.

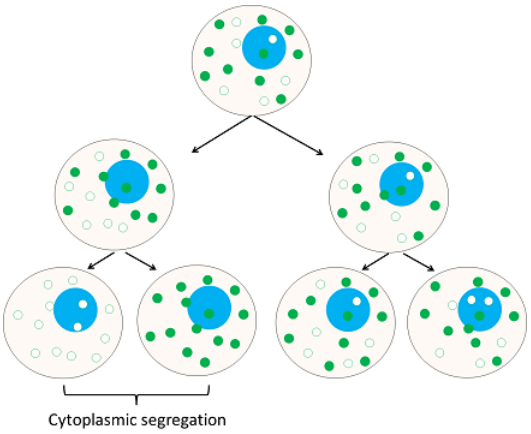


Figure 106. Cytoplasmic segregation

Green: wild-type chloroplasts; White: mutant chloroplast; Blue: nucleus.

Extranuclear inheritance in fungi

Poky Neurospora: Maternal inheritance can be clearly demonstrated in the poky mutant of the haploid fungus, *Neurospora crassa*. The mutant called poky has a slow growth phenotype due to a mitochondrial mutation. When two strains of the fungus are crossed, the maternal parent contributes the bulk of the cytoplasm to the progeny. If this maternal parent carries mitochondria containing the poky mutation, all the progeny will be poky (represented by brown shading in Figure 107.).

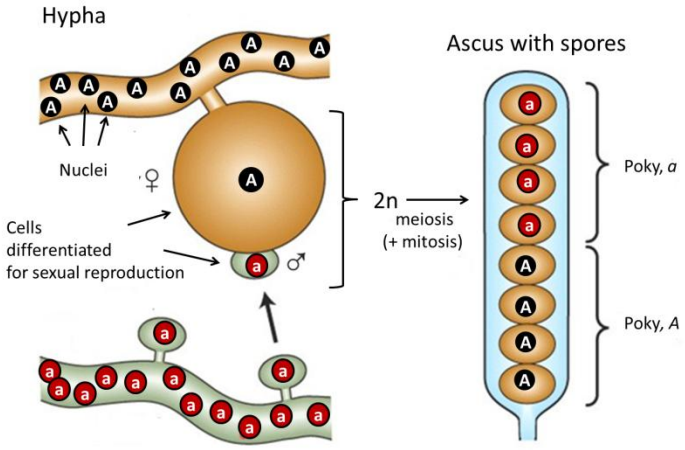


Figure 107. Inheritance of poky

The figure represents a cross between a poky mutant (brown shading) and a normal (grey shading) *Neurospora* strain.

In a reciprocal cross, where the maternal parent is normal, all the progeny would show normal growth. For a comparison, a nuclear gene with alleles "*a*" and "*A*" is shown, to illustrate the segregation of nuclear genes. This shows a 1:1 Mendelian ratio.

Petite mutants in yeast: Genetic studies on baker's yeast (*Saccharomyces cerevisiae*) have contributed greatly to the understanding of mitochondrial function and biogenesis. The wild-type cells of this yeast form relatively large colonies on solid medium. Scientists found a mutant that formed ***"petite" or small colonies***. Detailed analysis revealed that petite mutants were of several different types.

The ***cytoplasmic petite*** mutants ***showed extranuclear inheritance***, and contained large ***deletions in their mitochondrial DNA***. Some of these mutants lacked mtDNA altogether. In cytoplasmic petite mutants protein synthesis was abolished, and they had a deficient electron-transport chain. As a consequence, they had to rely on fermentation to produce ATP, and grow slowly → forming petite colonies.

Cytoplasmic male sterility (CMS)

Plasmids are small, extrachromosomal DNA molecules in the cell. They are common in bacteria, but also encountered in eukaryotes. Most eukaryotic plasmids are located in the mitochondria, and show inheritance mechanisms similar to those of mitochondria.

A well-studied example of ***mitochondrial plasmids*** is associated with ***cytoplasmic male sterility in maize***. Male sterility in plants is of great importance in agriculture. Male sterile plants produce no functional pollen. In hybrid maze production, using cytoplasmic male sterile plant as a female parent eliminate the need for detasseling (Figure 108.).

The mitochondrial male-sterile factor causes the abortion of pollen only if a ***nuclear gene "rf"*** (restorer of fertility) is in homozygous recessive form: *rf/rf*. A dominant allele of this gene in the nucleus of a male-sterile plant represses male-sterility and restores the fertility of the plant.

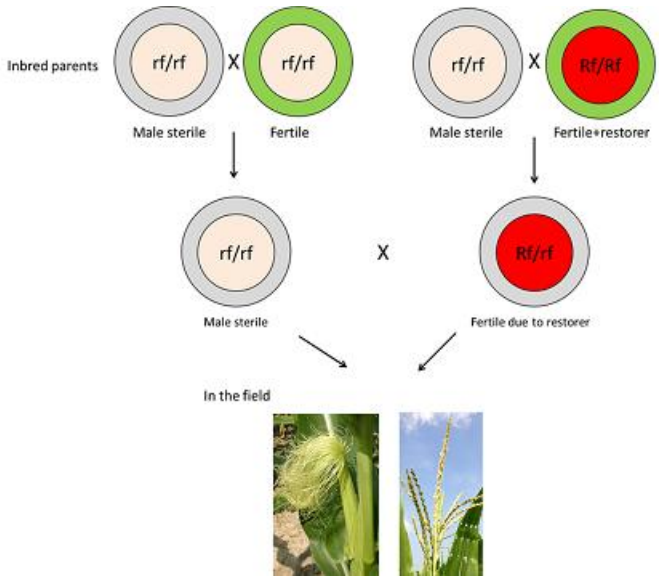


Figure 108. Production of hybrid corn using cytoplasmic male sterility

The hybrid corn is generated from four pure parental lines. Such hybrids are called double-cross hybrids. At each step, appropriate combinations of cytoplasmic genes and nuclear restorer genes ensure that the female parents will not self, and that male parents will have fertile pollen. Double-cross seed is produced in an isolated field, sold to the farmer, and planted to produce high-yielding maize.

6. Patterns of Chromosomal Inheritance

6.1 Autosomal inheritance

Mendel's experimental system

First, let us examine the pattern of inheritance for genes located on the ***autosomal chromosomes.*** (Autosomal chromosomes or ***autosomes*** are all the chromosomes in the nucleus except sex chromosomes.) The pattern of autosomal inheritance was first described by Gregor Mendel. Mendel's working method shows us an outstanding example of good scientific technique.

Mendel studied the ***garden pea*** (*Pisum sativum*). His choice had several advantages, the most important of which were that peas were available in a variety of shape and color that could be easily identified, and this plant can either self-pollinate (self) or cross-pollinate (cross). Other practical reasons were that peas were inexpensive, took up little space, had a short generation time and produced many offspring.

Mendel studied seven different characters (Table VI.). First he obtained ***pure lines*** for each of the characters he chose. (In breeding, these plants showed no variation in the character studied; all offspring produced by selfing or crossing were identical.) The differing forms can be called ***character forms, variants, or phenotypes***.

Character	Dominant form	Recessive form
Shape of ripe seeds	round	wrinkled
Color of seed interiors	yellow	green
Color of petals	purple	white
Shape of ripe pods	inflated	pinched
Color of unripe pods	green	yellow
Position of flowers	axial	terminal
Length of stems	long	short

Table VI. The seven Mendelian characters

Plants differing in one character / Monohybrid cross

Mendel pollinated purple flowered plants with pollen from white-flowered plants. He also made a ***reciprocal cross***: he transferred the pollen of purple-flowered plants onto the stigmas of white-flowered plants (Figure 109.).

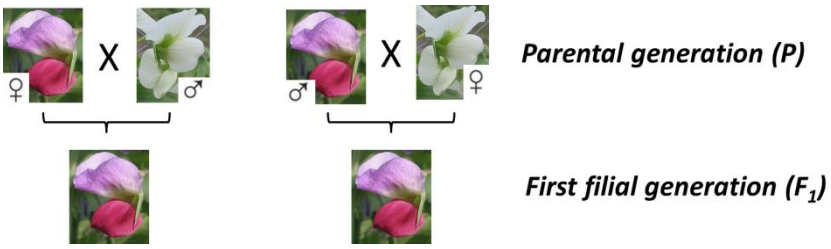


Figure 109. Crosses between pure-lined pea plants

The crosses produced the same result: all the plants in the first filial generation had purple flower. ***The first filial generation is uniform; it shows either of the two parental phenotypes.***

The second filial generation was made by self-pollination of the F_1 individuals. ***Mendel*** observed the F_2 individuals and ***quantified his results***. Mendel obtained 929 seeds from the selfing of F_1 individuals and planted them. In the second filial generation, he counted 705 purple-flowered and 224 white-flowered plants. The white phenotype, which was not observed in the F_1, reappeared in the F_2 generation. ***The original parental trait has been retained in the F_1 generation.***

Mendel described the "purple" phenotype ***which was expressed in the F_1*** generation as ***dominant***, and the "white" phenotype which ***was invisible*** as ***recessive***. ***The parental phenotype that is expressed in F_1 individuals obtained from crossing two pure lines is by definition the dominant phenotype.***

Mendel noted that the ratio of 705:224 is very close to a 3:1 ratio. He repeated the crossing experiments for the other 6 character pairs, and found the same 3:1 ratio in the F_2 generation in each experiment. How can this 3:1 ratio be explained? Mendel went on to thoroughly test those F_2 individuals that showed the dominant phenotype. In these experiments, he was working with the two phenotypes of seed color (Figure 110.). (The color of the seed is determined by the genetic constitution of the seed itself. This is convenient, because the phenotype of the seeds can be observed directly, without having to grow a plant from them.)

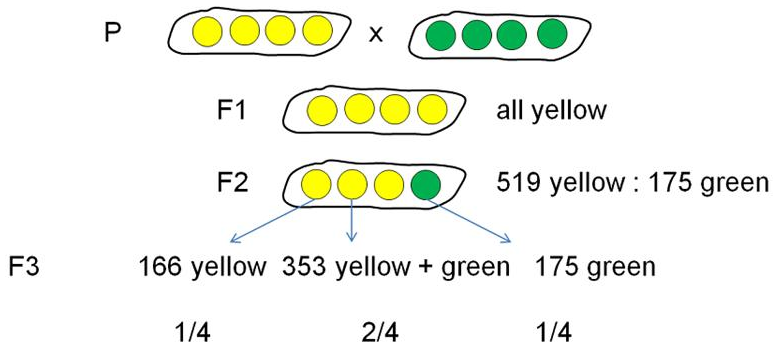


Figure 110. Testing the 3:1 phenotypic ratio with producing the F_3 generation

He crossed a pure yellow line with a pure green line and obtained an all-yellow F_1 generation. Therefore, yellow is the dominant and green is the recessive phenotype. Mendel grew F_1 plants from the seeds and then selfed the plants. In the F_2 generation, three-fourth of the seeds was yellow, and one-fourth was green. To further test the F_2 generation, Mendel selfed many individual plants. He found that the offspring of the green F_2 individuals were uniformly green, but there were two categories of F_2 yellow seeds. Some yellow F_2 individuals bore only yellow peas; others produced yellow and green peas, again in a 3:1 ratio. ***Underlying the 3:1 phenotypic ratio in the F_2 generation, there was a more fundamental 1:2:1 ratio.***

To explain his experimental results, Mendel devised a creative model:
1. He attributed the contrasting phenotypic differences and their inheritance patterns to certain ***hereditary particles***, which we now call ***genes***.
2. The "hereditary particles" have different forms, each corresponding to an alternative phenotype or character. These ***variants of genes*** we now call ***alleles***. Every individual carries two alleles of a gene, which can be the same or different. ***F_1 plants must have two different alleles***, one responsible for the dominant, and the other for the recessive phenotype.

3. When gametes are formed, the members of the allele pair separate; each gamete carries only one allele.
4. The members of the gene pairs (alleles) segregate (separate) equally into the gametes. 50% of the gametes carry one, and 50% of the gametes carry the other allele.
5. Upon fertilization, gametes unite randomly, that is, regardless of which allele they carry (Figure 111).

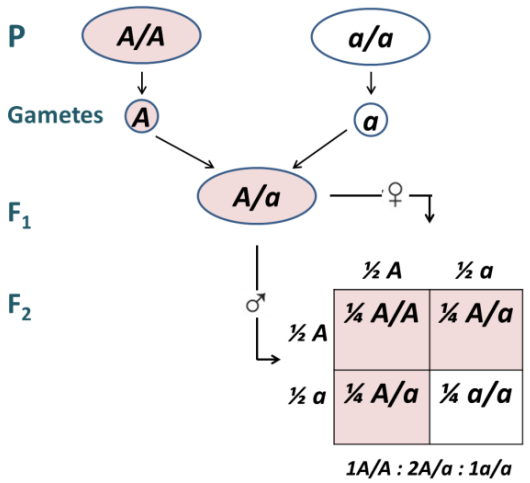


Figure 111. Explanation of the 1:2:1 ratio

Let "*A*" represent the allele that determines the dominant phenotype (e.g. purple flower or yellow seed), and "*a*" represent the allele responsible for the recessive phenotype (e.g. white flower or green seed). The members of the allele pair are separated by a slash (/).

A key part of the model is the nature of the F_1 individuals, which are supposed to carry two different alleles. If this is true, plants that have grown from yellow seeds must produce yellow and green progeny in a ***1:1 ratio*** when crossed with green plants. Mendel counted 58 yellow and 52 green plants. ***A cross of an individual organism of unknown genotype with a partner of recessive phenotype is called a testcross*** (Figure 112.).

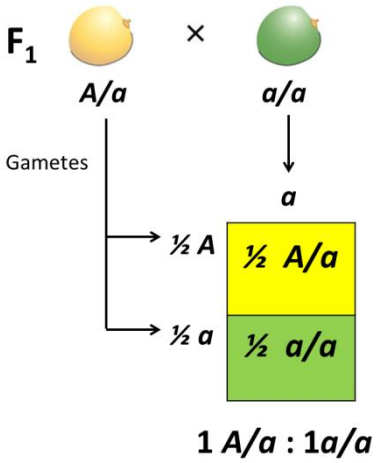


Figure 112. Testcross

Conclusions drawn from monohybrid crosses

Mendel's first law (the law of equal segregation): The two members of an allele pair segregate from each other into the gametes so that half the gametes carry one member of the pair and the other half of the gametes carry the other member of the pair.

Modern terms for Mendel's experiments:

"A" and "a" are alleles (variants of a gene); "A" stands for the dominant, and "a" stands for the recessive genotype.

An A/A individual is homozygous dominant; A/a is a heterozygous (hybrid), and a/a is a homozygous recessive individual.

"Pure breeding" plants are always homozygous.

A phenotype is a manifested trait, such as the purple or the white flower color.

In a monohybrid cross, there is a 1:2:1 ratio of the genotypes behind the 3:1 ratio of the phenotypes.

(Because individuals of the F_1 generation are heterozygous for one character, the experiment where the segregation of the allele pair of one gene is followed is called a monohybrid cross.)

Plants differing in two characters / Dihybrid cross

The two characters Mendel examined were seed color and seed shape. He started with two parental pure lines:

$$y/y \cdot R/R \text{ x } Y/Y \cdot r/r$$

where "*Y*"stands for yellow seed color and "*y*" stands for green seed color;
and "*R*" stands for round seed shape and "*r*" stands for wrinkled seed shape.
(A dot between the genes indicates that we do not know whether they are on the same chromosome or they are on different chromosomes.)

P:	$y/y \cdot R/R$ x $Y/Y \cdot r/r$
F_1:	$Y/y \cdot R/r$

The dihybrid F_1 seeds were all yellow and round. This showed that the dominance of *Y* over *y* and *R* over *r* was unaffected by the condition of the other gene pair in the dihybrid individuals. To obtain the F_2 generation, Mendel self-pollinated the F_1 plants. In the F2 generation, he counted 556 seeds, in the following distribution:

315 yellow, round
108 green, round
101 yellow, wrinkled
32 green, wrinkled

The F2 seeds were of four different types, in a 9:3:3:1 ratio. Mendel added up the number of individuals in certain F_2 phenotypic classes, to check if the monohybrid 3:1 F_2 ratios were still present:

Yellow: 315 + 101 = 416
Green: 108 + 32 = 140
Ratio: 2.97 : 1

Round: 315 + 108 = 423
Wrinkled: 101 + 32 = 133
Ratio: 3.18 : 1

Mendel realized, that ***the 9:3:3:1 ratio is the random combination of two 3:1 ratios***. The random combination of two 3:1 ratios can be visualized with a branch diagram (Figure 113.).

Figure 113. Branch diagram showing the combination of four phenotypes of two traits

For this situation, the product rule of matemathics applies: the probability of two independent events occurring simultaneously is the product of the individual probabilities. The probabilities of the combined phenotypic classes indicate that the ***genes that determine the shape and the color of the seed are inhereted independently***.

An F_1 dihybrid individual produces gametes of four different types, in a 1:1:1:1 ratio: *RY*, *Ry*, *rY*,and *ry*. Both the male and female gametes show the same proportions. The four female gametic types are fertilized randomly by the four male gametic types to obtain the F_2 generation. An illustrative representation of this can be done by using a 4x4 grid called ***Punnett square*** (Figure 114.).

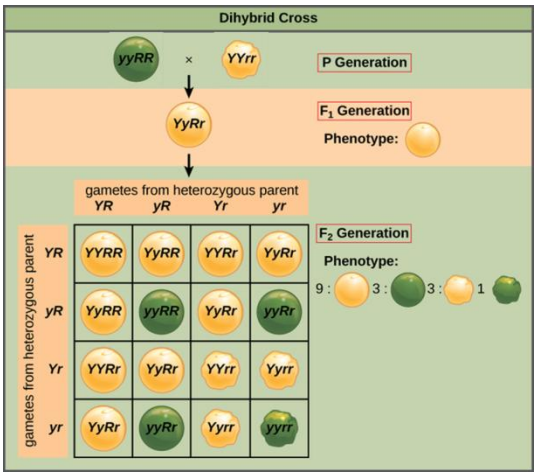


Figure 114. Punnett square for a dihybrid cross

Based on: https://commons.wikimedia.org/wiki/File:Figure_12_03_02.png
CNX OpenStax, CC BY 4.0 <https://creativecommons.org/licenses/by/4.0>, via Wikimedia Commons

Conclusions drawn from dihybrid crosses

Mendel's second law (the law of independent assortment): different allele pairs assort independently in gamete formation.

Mendel had no knowledge of the chromosomal location of the genes he studied. We know that most cases of independence are observed for genes that reside on different chromosomes.

Therefore, the modern version of Mendel's second law can be stated as follows: allele pairs on different chromosome pairs assort independently at meiosis.

Mendel tested his principle of independent assortment by crossing dihybrid F_1 individuals with fully recessive homozygotes (***test cross***):

Parents:		*Y/y;R/r*	x	*y/y*; *r/r*			
Gametes:		*YR, yR, Yr, yr*		only *yr*			
F_2:	*YyRr*		*yyRr*		*Yyrr*		*yyrr*
Ratios:	**1**	**:**	**1**	**:**	**1**	**:**	**1**

The progeny proportions in this cross are direct manifestation of the gametic proportions produced by the dihybrid.

6.2 Sex-linked inheritance

Cytogeneticists divide the X and Y chromosomes into ***homologous*** and ***differential regions***. For the location of homologous regions, also called pseudoautosomal regions on the human sex chromosomes, see Figure 32. on page 26. The X chromosome carries many genes, most of which are not involved in sexual functions, and have no counterpart on the Y chromosome. The Y chromosome contains only a few dozen genes. An important region of chromosome Y, the ***SRY (sex-determining region Y)***, determines maleness itself.

Genes of the homologous regions are substantially similar on both sex chromosomes. The differential regions contain genes that have no counterpart on the other sex chromosome, hence males are said to be ***hemizygous*** for these genes.

Genes on the differential region of the X show an inheritance pattern called ***X linkage***; those on the differential region of the Y show ***Y linkage***. In general, genes on the differential regions are said to show ***sex linkage***.

A gene that is sex-linked shows different patterns of inheritance from those of the genes on the autosomes. In autosomal inheritance, reciprocal crosses result in the same phenotypic proportions for the male and female progeny. In contrast, reciprocal crosses often produce male and female progeny that show different phenotypic ratios when the trait is sex-linked. In fact, this is a diagnostic sign of sex-linkage.

X-linked inheritance

The first example of sex-linked inheritance was described by Morgan, in *Drosophila*. The wild-type eye color of *Drosophila* is red. Morgan also had pure lines with white eye. This phenotypic difference is determined by a gene located on the differential region of the X chromosome. In *Drosophila*, genes are named after the first mutant allele found. Hence, "***w***" stands for the mutant allele, responsible for the ***white*** eye-color, and "$\boldsymbol{w^+}$" designate the wild-type allele, responsible for the red eye-color.

Morgan performed ***reciprocal crosses between red-eyed and white-eyed Drosophila*** that gave different results. The alleles are X-linked, and the ***inheritance of the X chromosome explains the phenotypic ratios*** observed (Figure 115.).

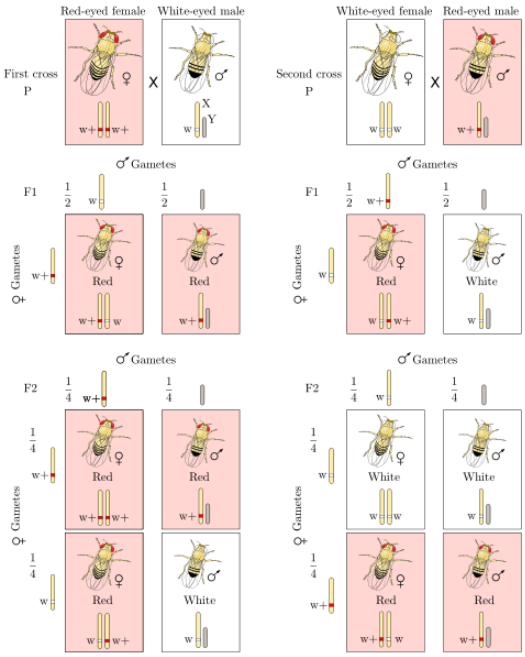


Figure 115. Reciprocal crosses between red-eyed and white-eyed *Drosophila*

https://commons.wikimedia.org/wiki/File:Sex-linked_inheritance.svg
GYassineMrabetTalk ✉ This W3C-unspecified vector image was created with Inkscape., CC BY-SA 3.0 <https://creativecommons.org/licenses/by-sa/3.0>, via Wikimedia Commons

6.3 Interactions between the alleles of one gene

Incomplete dominance / Intermediate inheritance

In Mendel's experiments, the phenotype of the homozygous dominant individuals (*A/A*) was indistinguishable from the phenotype of heterozygotes (*A/a*). This case of dominance is called ***full or complete dominance***. However, two different alleles in a heterozygote may produce several different phenotypes, and may cause deviations from the Mendelian ratios.

In case of ***incomplete dominance***, the phenotype of the heterozygous individual is distinct from and often intermediate to the phenotypes of the two homozygous individuals. This situation was formerly called ***intermediate inheritance***. This type of incomplete dominance was found in the four o'clock plant (***Mirabilis jalapa***), wherein pink color is produced when true-breeding parents of red and white flowers are crossed (Figure 116.).

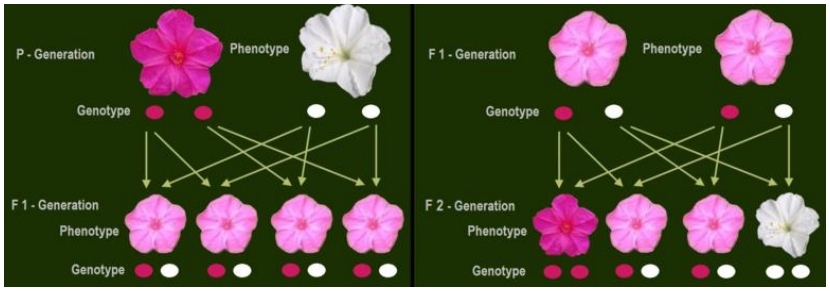


Figure 116. Inheritance of petal color in *Mirabilis jalapa*

Based on: https://commons.wikimedia.org/wiki/File:Intermediate_inheritance_-_incomplete_dominance.png
Sciencia58, CC0, via Wikimedia Commons

The ***1:2:1 ratio in the F_2*** generation shows that ***two alleles of a single gene*** produced these phenotypes. However, there are ***three different phenotypes***, and the phenotype of the ***heterozygous individuals*** is ***intermediate***. Crosses between red-flowered and white-flowered snapdragons (*Antirrhinum majus*) give similar results.

Codominance

A classical example of codominance is the ***inheritance of the MN blood groups in humans*** (Figure 117.). Blood groups are determined by an immunological antigen on the surface of the red blood cells. The phenotypic ratios correspond to the Mendelian inheritance, but there are three distinct phenotypes. Heterozygous individuals carry both antigens (antigen "M" and antigen "N") on their red blood cells.

Parents	Progeny phenotypes		
	MM	MN	NN
MM x MM	1	-	-
MM x MN	1/2	1/2	-
MM x NN	-	1	-
MN x MN	**1/4**	**1/2**	**1/4**
MN x NN	-	1/2	1/2
NN x NN	-	-	1

Figure 117. Inheritance of the MN blood groups

Two alleles are said to be codominant if the heterozygote shows both phenotypes that can be seen in the two homozygotes. In case of codominance, the genotype of an individual can directly be inferred from its phenotype.

Multiple alleles

So far we examined traits that are determined by two alleles of a gene. Although a diploid individual can possess only two alleles, several alleles of a single gene may be present in a population. The existence of these many alleles is called ***multiple allelism***; the set of alleles is called an ***allelic series***. Different relationships of dominance may exist between the different alleles.

AB0 blood group in humans: The four blood types in the AB0 system are determined by an ***allelic series of three alleles: i, I^A, and I^B*** (Figure 118.). The alleles I^A and I^B each determine a unique antigen on the surface of red blood cells, while the allele *i* confers inability to produce an antigen. The alleles I^A and I^B are both fully dominant to *i*. I^A and I^B are codominant.

Genotype	Blood group
I^A/I^A és I^A/i	A
I^B/I^B és I^B/i	B
I^A/I^B	AB
i/i	0

Figure 118. The AB0 blood group system

C gene in rabbits: Coat color in rabbits is determined by the ***allelic series of the C gene.*** The allele *C* confers full color; the coat is dark grey. The coat color of chinchilla (c^{ch}) rabbits is light greyish. Himalayan animals (c^{h}) are similar to the albinos, with black extremities. The *c* allele in homozygous form confers albino phenotype. The order of dominance in this allelic series can be described as follows: $C > c^{ch} > c^{h} > c$.

Clover chevrons: The white clover (*Trifolium repens*) is one of the most widely cultivated types of clovers. It shows considerable variation among individuals in the "V" or chevron pattern on the leaves (Figure 119.). In this species an allelic series determines the different chevron forms.

Figure 119. Chevron patterns

Two examples of chevron pattern on the leaves of two exceptional, four-leaved and five-leaved plants.
https://commons.wikimedia.org/wiki/File:Five-leaf_Clover,_Megan_McCarty128.jpg
Megan McCarty, CC BY 3.0 <https://creativecommons.org/licenses/by/3.0>, via Wikimedia Commons
https://commons.wikimedia.org/wiki/File:4-leaf_clover.JPG
Joe Papp, CC BY-SA 3.0 <https://creativecommons.org/licenses/by-sa/3.0>, via Wikimedia Commons

Incompatibility alleles in plants: Self-incompatibility is a well-known phenomenon in plants. Some plants produce both male and female gamete, but the fusion of these gametes never produces seeds. However, these plants are not sterile; they cross with certain other plants. Self-incompatibility has a genetic basis. In one of the most common self-incompatibility systems, ***one gene, S, determines compatibility-incompatibility relations*** (Figure 120.). This system was found, for example, in sweet cherries, tobacco, petunia and evening primrose.

The system is a nice example of multiple allelism. Those pollen grains, which bear an *S* allele that is also present in the maternal parent, cannot grow. Those pollen grains, which have different allele from the maternal parent, are efficient in fertilization. The number of *S* alleles in a series can be very large, which promotes outbreeding efficiently.

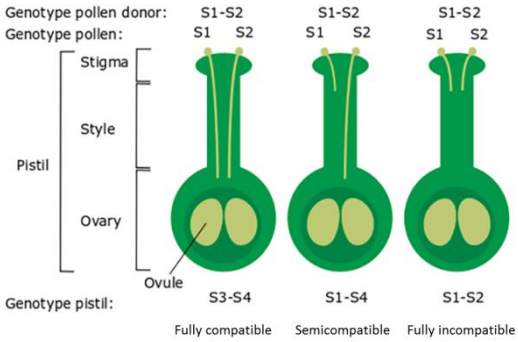


Figure 120. Example of a compatibility-incompatibility system

Based on: Figure 1, in Frontiers in Plant Science, April 2019, Volume 10, Article 407.

Lethal alleles

The normal coat color of mice is dark. Some mice have a lighter coat color called yellow. When a ***yellow mouse*** is mated to a normal mouse from a pure line, half of the progeny shows normal, the other half shows yellow coat color. From this observation one can conclude that: 1. a single gene determines these phenotypes; 2. the yellow mouse was heterozygous for this gene; 3. the "yellow" allele is dominant to the allele for normal coat color. However, crossing yellow individuals gave a surprising result (Figure 121.).

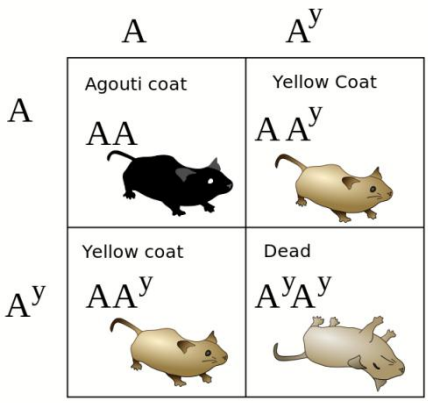


Figure 121. Punnett square of inheritance of lethal alleles

The figure shows the outcome of a cross between two mice with yellow phenotype.
https://commons.wikimedia.org/wiki/File:Lethal_alleles_punnett_square.svg
Jcfidy, CC BY-SA 4.0 <https://creativecommons.org/licenses/by-sa/4.0>, via Wikimedia Commons

Experimenters could not get a pure breeding line of yellow mice. Crossing heterozygous yellow mice resulted in a ***2:1 phenotypic ratio***, which is a departure from the Mendelian expectations. The homozygous form of the yellow allele (A^Y) is lethal. ***Lethal alleles in homozygous form cause the death of individuals which bear them.***

The yellow allele produces effects on two characters: coat color and survival. With respect to coat color, the yellow allele (A^Y) is dominant to the normal allele (A). However, with respect to viability, A^Y is recessive to A. Those genes that have more than one dictinct phenotypic effect are called ***pleiotropic genes***. Most of the genes are pleiotropic.

For another example of lethal alleles see the ***Manx cat***, in Figure 60., page 50. A single dose of the Manx allele, M^L, severely interferes with spinal development, resulting in the absence of tail in the heterozygotes. The homozygous form of this allele causes extreme developmental abnormality and death still in embryonic form.

6.4 Interacting genes and proteins

Complementation

Genes do not usually produce a phenotype on their own, but in interaction with other genes. Gene interactions may result in modified Mendelian ratios. An altered Mendelian ratio may give a hint to the type of interaction in question.

Let us examine the phenomenon of complementation with an example from harebell plants (genus *Campanula*). The wild-type flower of this plant is blue, but there are white-flowered variants as well. When two white-flowered plants are crossed, the F_1 individuals are blue, and the F_2 generation shows ratios typical of complementing genes (Figure 122.).

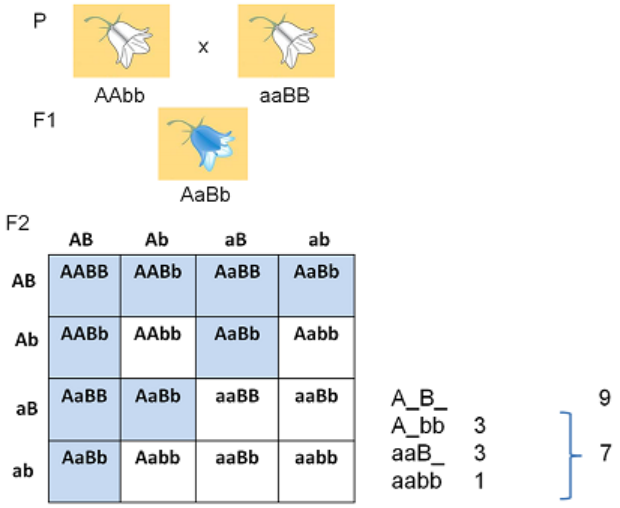


Figure 122. Complementation

The phenotypic ratio of complementary genes is 9:7.
When two haploid genomes bearing different recessive mutations are united in the same cell, and together they produce a wild-type phenotype, this phenomenon is called complementation.

Let us consider how complementation works at the molecular level. The blue color of the harebell flowers is provided by a pigment in the petals called anthocyanin. This pigment is made from chemical precursors that do not absorb light of any specific wavelength, and thus confer no color to the petals. In the synthesis pathway, each step is catalized by a specific enzyme encoded by a specific gene. A homozygous mutation in either of the genes prevents the completion of the pathway, and the accumulation of the corresponding precursor makes the plant white (Figure 123.).

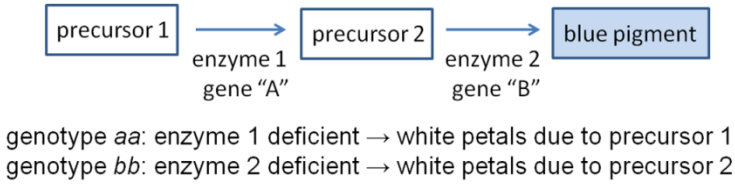


Figure 123. Molecular explanation of the petal colors of harebell flowers

The procedure called ***complementation analysis*** helps us to decide whether two recessive mutants with similar phenotypes carry a mutation in the same gene, or in different genes. Let us cross the two mutants with similar phenotype (e.g. two white-colored *Campanula*), and examine the F_1 generation (Figure 124.).

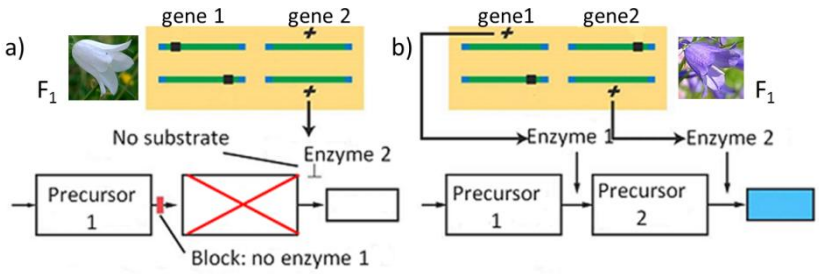


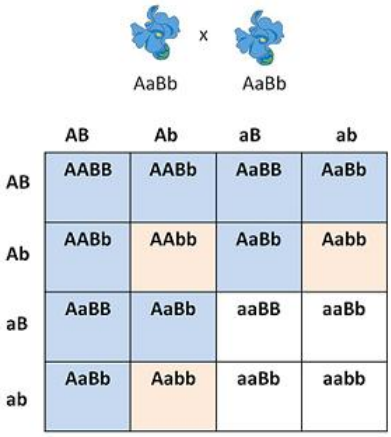
Figure 124. Complementation analysis

a) Mutation occurred in the same gene → no complementation → white-colored F_1 flowers.
b) Mutation in different genes → complementation → blue-colored F_1 flowers.

Epistasis

When an allele of one gene prevents the expression of an allele of another gene, masking its phenotype, this gene interaction is called epistasis. In the hierarchy of alleles, the one whose phenotype manifested is called ***epistatic***, and the allele whose phenotype is eliminated called ***hypostatic***. Epistasis is usually observed when the genes are involved in a developmental or biochemical pathway.

Recessive epistasis: The petals of the plant blue-eyed Mary (*Collinsia parviflora*) can be blue, white and magenta. When two pure lines (white and magenta) are crossed, the first generation is all blue. Selfing of the F_1 generation produces a ***9:3:4 phenotypic ratio*** in the F_2 generation (Figure 125.).



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

Figure 125. Punnett square of the F_2 generation showing recessive epistasis

The recessive allele of the "A" gene (aa) masks the expression of "B". In case of the plant *Collinsia parviflora*, the order of genes in a biochemical pathway determines the epistatic relationship.

Another case of recessive epistasis is the coat color of Labrador retriever dogs. The coat color can be black, brown or yellow. The alleles of one gene produces black (allele "*A*") or brown (allele "*a*") coat color. The recessive allele of another gene ("*bb*") is epistatic on these alleles, providing yellow color.

Therefore, the genotypes *A/_; bb* and *a/a ; bb* produce yellow coat color, while *A/_; B/_* and *a/a; B/_* black and brown, respectively. In this case the epistatic gene is developmentally

dowstream to the gene responsible for pigment production. Yellow dogs can produce black or brown pigments, but the pigment is not deposited in hairs.

Dominant epistasis: Three genes interact to determine the petal color of foxglowes (*Digitalis purpurea*; Figure 126.). "*C*" determines the ability to produce reddish pigment. The genotype *cc* results in white petals. The second, modifier gene determines the amount of pigment produced. "*B*" determines the synthesis of large amounts of pigment, resulting in dark red petals, while plants of *bb* genotype produce low amounts of pigment and have light reddish petals. The third gene affects pigment deposition. "*A*" prevents pigment deposition in the petal except in the throat spots, whereas the genotype *aa* allows pigment deposition all over the petal.

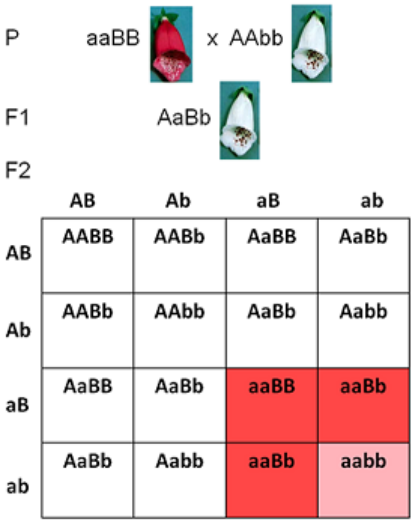


Figure 126. Dominant epistasis

A dark reddish (*CCaaBB*) and a white (*CCAAbb*) plant are crossed, and the F_1 generation is propagated further. (*CC* is omitted from the pedigree because it is homozygous in every individual.)

The dominant allele A prevents the expression of phenotypes conferred by the B and b alleles. The phenotypic ratio of dominant epistasis is 12:3:1.

Duplicate genes

The fruit shape of the plant sepherd's purse (*Capsella bursa-pastoris*) can be "heart-shaped" or narrow (Figure 127.).

Crossing plants of two different phenotypes results in a "heart-shaped" F_1 generation.

The characteristic phenotypic ratio in F_2 is 15:1.

It is easy to understand the molecular basis of this phenotypic ratio. In a biochemical pathway, the same step is catalyzed by two enzymes. The presence of either (that is, one dominant allele) is sufficient for product formation.

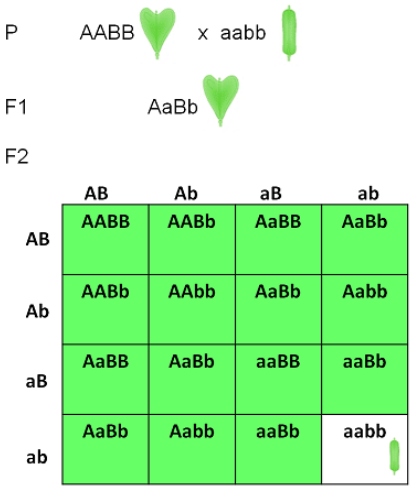


Figure 127. Duplicate genes

Penetrance and expressivity

Genes do not work in isolation but interact with each other, and their expression is influenced by the environment, as well. Therefore, the phenotype expected by the genetic constitution of individuals may not appear at all, or may be observable to variable extent in different individuals (Figure 128.). The terms penetrance and expressivity describe these situations.

Penetrance is defined as the percentage of individuals with a given genotype who exhibit the phenotype associated with that genotype.

Expressivity measures the extent to which a given genotype is expressed in an individual at the phenotypic level.

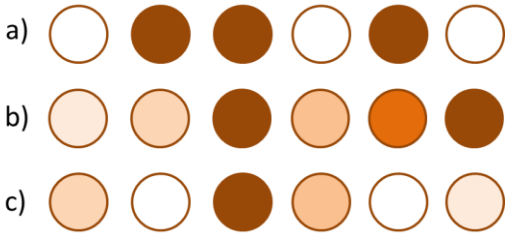


Figure 128. Expressivity and penetrance

Each circle represent an individual; the genetic constitution of the 18 individuals is identical, with respect to the gene studied. a) Variable penetrance: only 3 out of 6 individuals express the phenotype ("brownish color") expected by the genetic constitution. The gene for brown color has 50% penetrance. b) Variable expressivity: each individuals show the brownish color phenotype, to variable extent. The penetrance here is 100%.
c) Variable penetrance and expressivity.

7. Basic Terms of Population Genetics

7.1 Genetic variation in natural populations

The subject of population genetics

Organisms do not live as isolated individuals in nature, but interact with one another in groups called populations. Populations can be defined as groups of individuals of the same species that coexist at a specific space and time, and reproduce with each other. ***Factors that determine or affect the genetic composition of a population, and predictions about how that composition changes in time are the subject of population genetics.***

The genetic composition of a population can be described as the collection of frequencies of different genotypes. Processes that determine and change these frequencies act at the level of individual organisms. To relate the processes that work at individual-level to the genetic composition of populations, several phenomena must be considered:

1. Different ***mating patterns*** affect the genotypic composition of the population.
2. ***Migration*** of individuals may change the size and composition of the population.
3. ***Mutations*** may introduce new genetic variation into the population.
4. New combination of characters may arise by ***recombination***.
5. ***Natural selection*** may cause changes in population composition.
6. The consequences of ***random fluctuations*** in the actual reproductive rates of different genotypes should also be considered.

Population genetics analyze the effect of these factors on the genetic composition of populations.

Studying variation

When studying natural populations, the first step is the description of genetic variation among the individuals in the population. The complexity of the relationship between phenotype and genotype is very variable, depending on the character observed. At one extreme, there is a perfect one-to-one correspondence between genotype and phenotype. At the other extreme, several genes as well as environmental effects contribute to the manifestation of a certain phenotype. Quantitative genetics deals with studying such characters. For these characters, it is not possible to make precise statements about the undelying genotypic variation.

Population genetics has concentrated on ***characters with simple relations to the genotype***. In this case, the ***different phenotypes are the result of different allelic forms of a single gene***. A good example for this is the MN blood group system in humans (Table VII.).

	Genotype			*Allele frequencies*	
Population	*M/M*	*M/N*	*N/N*	*p(M)*	*q(N)*
Eskimo	0.835	0.156	0.009	0.913	0.087
Australian Aborigine	0.024	0.304	0.672	0.176	0.824
German	0.297	0.507	0.196	0.550	0.450

Source: W. C. Boyd, *Genetics and the Races of Man*. D. C. Heath, 1950.

Table VII. Genotype and allele frequencies in the MN blood group system in some human populations

In studying variation, first the phenotypic variation is described. Then, the underlying genetic variation is determined. As there is a perfect correspondence between genotype and phenotype in case of the MN blood group system, now these two steps merge into one. However, in most cases additional experiments or observations are needed to uncover the exact underlying genotype of each phenotype.
The simplest description of a single-gene variation is the determination of ***genotype frequencies***. It is more typical, however, to calculate ***allele frequencies***.

Calculation of allele frequencies: In a population, $f_{A/A}$, $f_{A/a}$ and $f_{a/a}$ are the frequencies of genotypes *A/A*, *A/a*, and *a/a*, respectively. Let "p" stand for the frequency of allele *A*, and "q" for the frequency of allele *a*. Because each homozygous *A/A* individual carries two, whereas each heterozygous *A/a* individual carries one *A* allele, the frequency of allele *A* can be calculated as follows:

$$\mathbf{p = f_{A/A} + ½\ f_{A/a}}$$

For the calculation of the frequency of allele *a*, the same logic applies:

$$\mathbf{q = f_{a/a} + ½\ f_{A/a}}$$

If there are only two alleles in the population:

$$\mathbf{p + q = f_{A/A} + f_{A/a} + f_{a/a} = 1,}$$
and $\mathbf{q = 1 - p}$

If there are more than two different alleles, for the calculation of each allele frequencies, the frequency of the homozygous form of a certain allele, plus the sum of half of the frequencies of all heterozygous forms where the allele of question occurs must be calculated.

Some examples of polymorphism

In natural populations, usually more than one form of a gene and more than one phenotype for a character occurs (Figure 129.). Considerable genetic variation, or ***polymorphism***, can be observed within populations and between populations as well.
Sometimes one form of the gene or character is preponderant in the population, and only some exceptional individuals carry an unusual variant. In this case, the very common form is called the ***wild-type***, in contrast to the rare mutants. In other cases, there are several common forms.
Genetic variation, which can be the basis for evolutionary changes, ***is ubiquitous and observable at different levels.***

Figure 129. Representative shell colors of the snail *Littorina saxatilis*
(Source: Figure 1 in Heredity (2017) 118, 21-30.)

Examples of protein polymorphism: ***Immunological polymorphism*** is a widespread phenomenon in the living world. For example, in humans more than 40 blood group systems are described according to different antigens on the surface of red blood cells. One of the most important ones is the ***AB0 system***, where two alleles of three different types (I^A, I^B, i) determine the blood group of each individual. The frequencies of different alleles vary considerably among different human populations (Table VIII.).

Population	I^A	I^B	i
Eskimo	0.333	0.026	0.641
Sioux	0.035	0.010	0.955
Japanese	0.279	0.172	0.549

Source: W. C. Boyd, *Genetics and the Races of Man*. D. C. Heath, 1950.

Table VIII. Allele frequencies in the AB0 blood group system in some human populations

Mutations in the coding region of a gene may lead lo changes in the amino acid sequence of the encoded protein. The exact nature of these changes can be revealed by sequencing the gene. However, this is a time-consuming and laborious process. It is frequently enough to detect different variants of a protein without knowing the particular changes and the exact amino acid sequence. Amino acid substitutions may cause a change in the charge of the protein, and such an alteration of the physical properties can be detected by protein gel electrophoresis (Figure 130.). The ***amino acid sequence polymorphism*** is very frequent in populations. Studying a very large number of species, researchers have found that generally about one third of the genes coding for proteins are polymorphic at the protein level, and about 10% of an individual's genome is in heterozygous state.

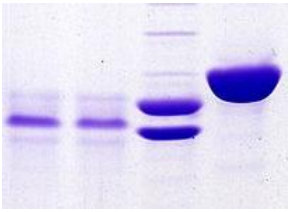


Figure 130. Portion of a Coomassie blue-stained protein gel showing four samples

Examples of DNA polymorphism: "Karyotype" is generally defined as "the set of chromosomes, typical of a species". For many species, to define such a standard set of chromosomes is difficult, because they show a large variety of chromosome number and morphology. Supernumerary chromosomes, reciprocal translocations and inversions can be observed in many population of plants, insects, and even mammals. An example of chromosomal polymorphism is presented below, in Table IX. Using molecular biological methods, DNA polymorphism at the finest (nucleotide) level can be revealed.

No supernumeraries or translocations	Translocations	Supernumeraries	Both translocations and supernumeraries
0.560	0.133	0.256	0.042

Source: H. Lewis, *Evolution* 5, 1951, 142-157.

Table IX. Chromosome polymorphism in a population of the plant *Clarkia elegans*

7.2 The effect of mating patterns on genetic variation

Random mating and the Hardy-Weinberg equilibrium

The genetic composition of natural populations is usually too complex to allow scientists a detailed examination. Applying a number of simplifications, population geneticists defined a ***model population*** or ***ideal population***.

This ideal population meets the following requirements:

- Within the population, males and females mate with one another completely at random.
- The population size is large, so genetic drift cannot occur.
- Allele frequencies in the population are not influenced by mutation or migration.
- There is no selection, that is, parents of different genotypes are equally efficient in producing progeny.

The ideal population has a single gene locus with a dominant allele "*A*", and a recessive allele "*a*". Such a population consists of three genotypes: *A/A*, *A/a*, and *a/a*. Let us examine the effect of random mating on the allele frequencies of the next generation.

If, in this population, the allele frequency of *A* is 0.6 in both sperms and eggs, random fertilization results in 0.6 x 0.6 = 0.36, that is, 36 percent of offspring with a homozygous dominant phenotype *A/A*.

In the same way, the frequency of homozygous recessive offspring will be 0.4 x 0.4 = 0.16, that is, 16%.

A heterozygous *A/a* offspring can be the results of pairing an *A* egg with an *a* sperm with a probability of 0.6 x 0.4, or the reverse combination, having the same probability. Thus, the frequency of heterozygous offspring is 2 x 0.6 x 0.4 = 0.48, that is, 48%.

If we calculate allele frequencies in this new generation, it becomes obvious that ***random mating does not change allele frequencies.*** The proportions of homozygotes and heterozygotes remain the same in each successive generation.

These constant frequencies form the ***equilibrium distribution*** that can be calculated according to the formula

$$\underset{p^2}{A/A} + \underset{2pq}{A/a} + \underset{q^2}{a/a} = 1$$

where p stands for the frequency of allele *A*, q stands for the frequency of allele *a*, and p+q=1. This distribution is called the ***Hardy-Weinberg equilibrium*** (Figure 131.).

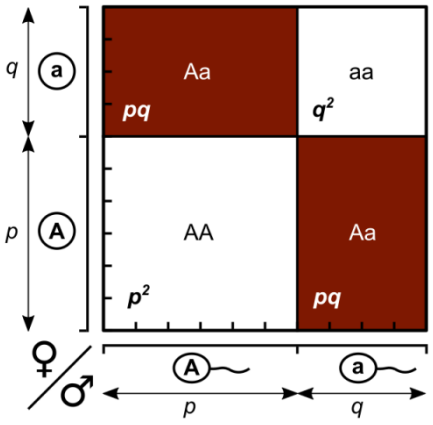


Figure 131. The Hardy-Weinberg equilibrium frequencies that result from random mating

https://commons.wikimedia.org/wiki/File:Hardy%E2%80%93Weinberg_law_-_Punnett_square.svg

The Hardy-Weinberg equilibrium means that ***sexual reproduction maintains genetic variation in the populations***, in the absence of disturbing forces. Further, it shows that irrespectively the particular genotypic composition of the parental population, the ***genotypic distribution of the next generation is completely specified by the allele frequencies***.
The presentation of genotype frequencies as a function of allele frequencies (Figure 132.) allows drawing a few conclusions.

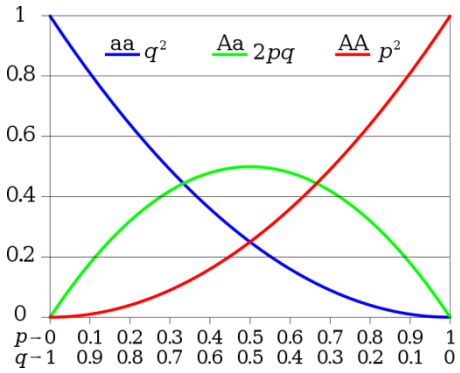


Figure 132. Homozygote and heterozygote frequencies as a function of allele frequencies

https://commons.wikimedia.org/wiki/File:Hardy-Weinberg.svg
Johnuniq, CC BY-SA 3.0 <https://creativecommons.org/licenses/by-sa/3.0>, via Wikimedia Commons

The ***measure of genetic variation*** is the amount of ***heterozygosity*** for a gene in a population. This heterozygosity can be directly determined by counting heterozygotes, or it can be calculated from the allele frequencies. For two alleles, the proportion of ***heterozygotes is the highest if the alleles occur at equal frequencies*** in the population. Generally, heterozygosity is the greatest when there are many alleles of a gene, all at equal frequency.

Another consequence of the Hardy-Weinberg rule is that ***rare alleles are virtually never in homozygous condition***. If the frequency of a disease-causing mutant allele is, e.g. q=0.01, the frequency of homozygotes is 0.0001, so the disease affects only 1 out of 10 000 individuals of the population.

Deviations from random mating

The ideal population is essentially ***panmictic***. In such a population, mating partners are chosen without any precondition or preference. However, real populations may deviate from random mating in two different ways.

Preferences for choosing a partner may be based on certain ***phenotypic*** features. This is typical of the assortative systems. Individuals may tend to choose each other because they resemble to each other in some traits. For example, in human there is a ***positive assortative mating*** bias for skin color and height. When dissimilar individuals tend to choose each other as mating partners, it is called ***negative assortative mating***.

The second type of deviation from random mating is based on the ***genotypic*** similarity or dissimilarity of the mating partners. If individuals that are relatives mate with each other more frequently than would occur by chance, then the population is ***inbreeding***. If mating between relatives is less common than would occur by chance, it is called ***negative inbreeding or enforced outbreeding.***
From the four possible mating patterns, ***inbreeding is especially important in agriculture***.

Inbreeding

When two individuals that are related mate, there is a chance that they pass on to their descencents exactly the same allele. This is because both parents inhereted this common allele from a common ancestor. Thus, to the chance of homozygosity (p^2+q^2) that arises from the random mating of unrelated individuals, an extra chance of ***homozygosity by descent*** is added (Figure 133.). The probability of homozygosity by descent is called the ***inbreeding coefficient*** (F).

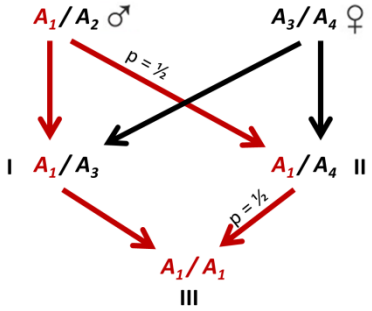


Figure 133. Calculation of homozygosity by descent
For the explanation of the figure, refer to the main text.

In the example shown in Figure 133, individuals I and II, who are full sibs, mate to produce individual III. Individual III has received one copy of allele A_1 from its grandfather through individual I. The chance that individual II receives A_1 from its father is ½, and if it does, the chance that II will pass A_1 on to individual III is also ½. Thus, the chance that individual III will be homozygous A_1/A_1 by descent from the original ancestor is ½ x ½ = ¼.

Inbreeding can have deleterious consequences. In the human populations, there are many deleteroius alleles for different genes. The chances are high that an offspring of close relatives will be homozygous for such an allele.

Inbreeding occurs at maximum rate when individuals in a population reproduce by ***self-fertilization*** (Figure 134.).

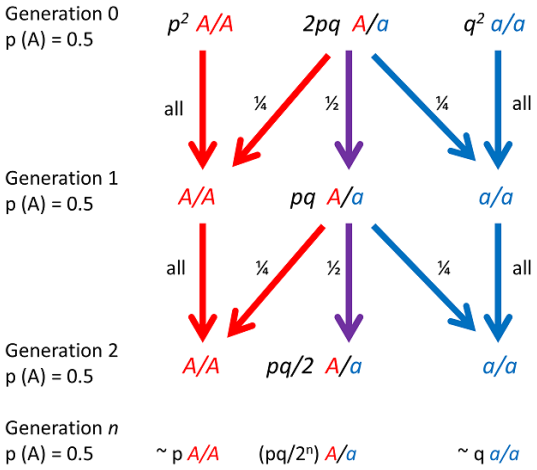


Figure 134. Repeated generations of inbreeding / self-fertilization
For the explanation of the figure, refer to the main text.

Suppose that two alleles at equal frequencies are present in such a population, that is p*(A)*=q*(a)*=0.5. The ratio of heterozygotes is reduced by 50% in every successive generation. This pattern is repeated until the frequencies of the resulting homozygous lines *A/A* and *a/a* become equal with the original allele frequencies.
Inbreeding converted the genetic variation observed *within* the original population into variation *between* homozygous inbred lines derived from this population.

Systematic inbreeding between close relatives leads to ***complete homozygosity of the population.*** Generally, from two alleles present in an inbreeding population one will eventually be lost, and the other will become ***fixed*** (have a frequency of 1.0).

Assortative mating systems

Both ***positive assortative mating*** and inbreeding ***increase the homozygosity in the population.*** However, there is an important difference in the two mating systems: assortative mating is specific to a particular phenotype, whereas inbreeding applies to the whole genome.

Plant populations that show a degree of variation in flowering time present a good example of positive assortative mating. In such populations, individuals that bloom at the same time produce progeny together more frequently, than it would be expected in a panmictic system.

In the negative assortative mating system, mating partners of dissimilar phenotypes prefer each other. This is typical of several plant species, because several mechanisms that prevent self-pollination and promote cross-pollination operate on this bases (Figure 135.). ***Negative assortative mating maintains heterozygosity for the definitive phenotype*** of the system.

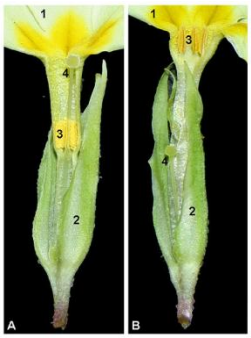


Figure 135. Heterostyly (distyly) in common primrose (*Primula vulgaris*)

1. Petal; 2. Sepal; 3. Anther; 4. Pistil. The two floral morphs are called L-morph (A) and S-morph (B). The pollen of the same flower cannot land on its own stigma in either morphs, thus distyly enhances allogamy.
https://commons.wikimedia.org/wiki/File:Distyly_primula.jpg
berru, CC BY-SA 3.0 <http://creativecommons.org/licenses/by-sa/3.0/>, via Wikimedia Commons

7.3 The sources of genetic variation

Mutation

Mutation, recombination, and migration can contribute to the genetic diversity of a population. However, ***the ultimate source of genetic variation is mutation, because only mutation can create new alleles***.

Spontaneous mutations occur very infrequently. ***Mutation rate (μ)*** is defined as the probability that an allelic form of a gene (*A*) changes into another allelic form (*a*) in one generation (Table X.).

Organism	Gene	Mutation rate per generation
Bacteriophage	Host range	2.5×10^{-9}
Escherichia coli	Phage resistance	2×10^{-8}
Zea mays (corn)	*R* (color factor)	2.9×10^{-4}
	Y (yellow seeds)	2×10^{-6}
Drosophila melanogaster	Average lethal	2.6×10^{-5}

Source: T. Dobzhansky, Genetics and the Origin of Species, 3rd ed., rev. Columbia University Press, 1951

Table X. Point-mutation rates in different organisms

The increase in the frequency of a mutant allele (*a*) depends on the mutation rate (μ) and the frequency of the nonmutant allele (*A*). Because μ is generally very small, and the number of nonmutant alleles is decreasing in every generation, ***the rate of increase of the mutant allele is extremly slow***, and it gets gradually slower with every generation. For example, if the constant mutation rate (μ) is 10^{-5}, only 10% mutant allele (*a*) is produced after 10000 generations. ***Due to the extremly slow mutation rates, mutations alone cannot be responsible for rapid genetic changes of populations***.

Recombination

Haplotype can be defined as the combination of alleles of different genes on the same chromosome that are inhereted together. In a double heterozygote *BC/bc* that inhereted *BC* and *bc* haplotype from his parents, recombination between the *B* gene and the *C* gene can produce new haplotypes: *Bc* and *bC* (see Figure 80, on page 59). Repeated recombination between genes tend to randomize combinations of alleles of different genes.

Recombination is a mandatory event of meiosis in the vast majority of ***sexually reproducing*** organisms. ***Recombination creates genetic variation usually much faster than mutation.*** Organisms that reproduce asexually do not have this source of variation, so the populations of asexual organisms may change more slowly than sexual organisms.

Migration

Genetic variation of a population can be increased by migration. If immigration occurs from a population that has different gene frequencies than the recipient population, the resulting new frequencies of the recipient population will be somewhere between the gene frequencies of the donor population and the original gene frequency of the recipient population.

The change in gene frequency depends on two factors: the actual difference in gene frequencies of the donor and recipient populations, and the proportion of the recipient population that is made up of new migrants. Suppose the frequency of allele "*A*" is 0.8 in the recipient, and 0.3 in the donor population, respectively, and the proportion of immigrants is 10% in the newly formed population. In this case, the frequency of *A* allele changes to 0.75 after the immigration (0.8x0.9+0.3x0.1). Unlike mutation, ***migration can change gene frequencies substantially, in shorter periods of time.***

7.4 Selection

Natural selection

Several factors have been considered so far that contribute to the genetic diversity of populations, but these factors alone cannot explain why and how organisms can survive in an ever-changing environment. A simple but powerful explanation to this question lies in ***Darwin's theory of natural selection***. The essence of his theory is that ***individuals best adapted to their environments are more likely to survive and reproduce*** (Figure 136.).

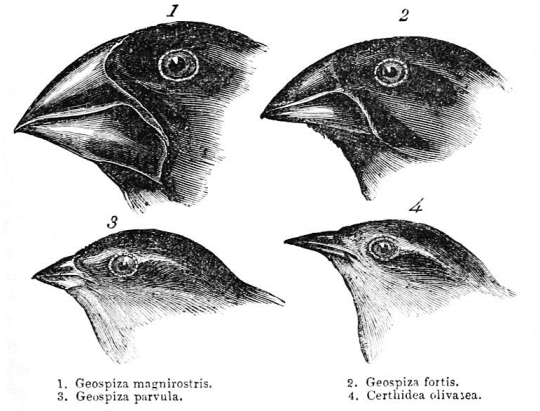


Figure 136. Four species of Darwin's finches

The beaks of these birds are variously adapted to feed on seeds, buds and insects.
https://commons.wikimedia.org/wiki/File:Darwin%27s_finches_by_Gould.jpg
John Gould (14.Sep.1804 - 3.Feb.1881), Public domain, via Wikimedia Commons

Natural selection means the differential rates of survival and reproduction of individuals of different phenotypes. Natural selection acts on phenotypes, because these characteristics of the organism interact with the environment. However, ultimately it is the genetic basis of a particular phenotype (the heritable genotype) that gives the phenotype a reproductive advantage. The process of ***selection alters the frequencies of different genotypes in the population*** by choosing the individuals with the most advantageous variation. ***Natural selection is a key mechanism of evolution***; populations that evolve to be sufficiently different may become different species eventually.

Fitness

Darwinian fitness is defined as the relative probability of survival and reproduction for a genotype. This term is not to be confused with the "physical fitness of a person", as it is meant in the everyday sense. First of all, the ***concept of fitness applies to the average*** probability of survival and the average reproductive rate of individuals ***in a genotypic class***.

Further, Darwinian fitness is ***affected by two groups of factors***: factors, that affect the survival chances of individuals, and factors that affect their efficient reproduction. During the first stage of life cycle, until individuals reach sexual maturity, fitness differences of the different genotypes are actually due to differences in viability. When individuals reach the reproductive stage in their life cycles, factors that affect their efficient reproduction become very important in determining their fitness.

Allelic substitution at a single locus rarely makes enough difference in the phenotype to measurably alter fitness. However, the "inborn errors of metabolism" make an exception:

homozygotes for those alleles that cause e.g. sickle-cell anemia (Figure 137.) or phenylketonuria show severely decreased fitness.

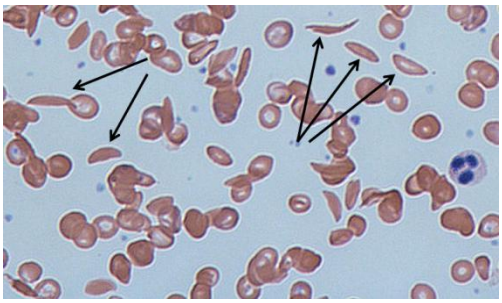


Figure 137. Blood sample of a person with sickle-cell disease

The basic problem is the abnormality of the β-globin protein, a subunit of the oxygen-carrying protein hemoglobin. A number of health problems occur in the affected indivuduals, ultimately leading to a shorthening of the life expectancy (Arrows point at some abnormal red blood cells.)
Based on: https://commons.wikimedia.org/wiki/File:Sickle_cells.jpg
Dr Graham Beards, CC BY-SA 3.0 <https://creativecommons.org/licenses/by-sa/3.0>, via Wikimedia Commons

Measuring fitness is relatively easy when the genotypes differ at many loci. In an experiment carried out with *Drosophila pseudoobscura*, several different lines sampled from natural environment were made homozygous for the second chromosome. Then, the probability of the survival of different lines from egg to adult was tested at three different temperatures. Some lines showed similar viabilities at the three different temperatures tested, but this was rather the exception than the rule. Most of the lines tested exhibited markedly different viabilities at the different temperatures; some showed better survival at the higher temperatures, but the viability of most lines decreased with the increasing temperature. This experiment demonstrated that ***the fitness*** (in this case viability, a component of fitness) ***is different in different environment.***

Selection alters allele frequencies

The simplest way to demonstrate the effect of selection on allele frequencies is to use a lethal allele as an example. Let us assume that allele *a* is lethal in homozygous form, causing individuals with *a/a* genotype to die before reaching reproductive age. If the frequency of *a* in a population is 0.2 (q=0.2), random mating will result in three genotypes with the proportions of

$$0.64\ A/A : 0.32\ A/a : 0.04\ a/a.$$

However, homozygotes *a/a* die, leaving only *A/A* and *A/a* individuals to reproduce. To determine the frequency of the lethal *a* allele among the gametes of survivors, first the proportions of *A/A* and *A/a* must be recalculated so that the total propotions of *A/A* and *A/a* add up to 1.00.

$$0.64/0.96 = 0.667\ A/A \quad \text{and} \quad 0.32/0.96 = 0.333\ a/a.$$

The frequency of the lethal *a* allele among the gametes of survivors is 0.333/2 = 0.167. During a single generation, the frequency of the normal *A* allele has increased, while the frequency of the lethal *a* allele has decreased by 0.033.

When each genotype has some relative probability of survival ($W_{A/A}$, $W_{A/a}$ and $W_{a/a}$), the following general formula can be applied for the calculation of changes in allele frequencies:

$$\Delta p = \frac{pq\,(\overline{W}_A - \overline{W}_a)}{\overline{W}}$$

where Δp stands for the change in allele frequency p; p and q stand for the frequency of *A* and *a* alleles, respectively;

$\overline{W}_A$ and $\overline{W}_a$ denote the mean fitness of allele *A* and *a*, respectively; and $\overline{W}$ means the mean fitness of the population.

Selection works in a way that the allele with the higher mean fitness increases in frequency in the population. As a consequence, the average fitness of the population increases as well, so selection can be best described as a process that ***increases mean fitness***.

The general expression for the change in allele frequency explains why selection works very slowly in a genetically uniform population. The speed of change depends not only on the difference in fitness between the alleles, but also on the factor *pq*. If the frequency of an allele is very small, then *pq* is nearly 0, and selection proceeds very slowly. ***Selection depends on genetic variation***.

Different forms of selection

Next, four different forms of selection, depicted in Figure 138, will be described briefly.

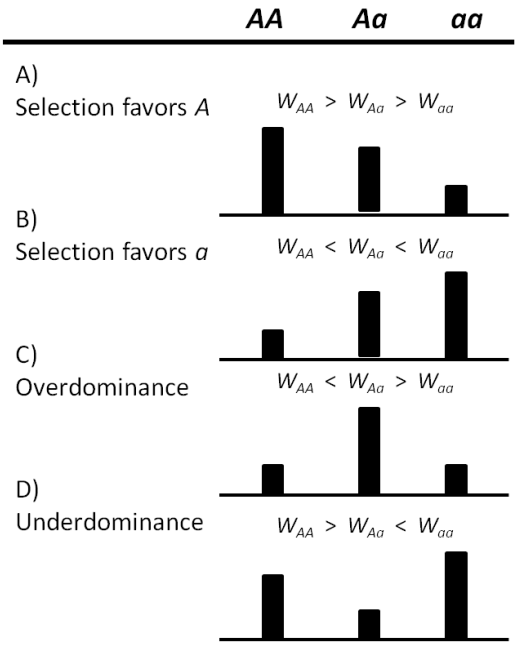


Figure 138. Forms of selection

A) When homozygotes that carry dominant allele are more fit than homozygotes that carry recessive allele, and the fitness of heterozygotes is somewhat in between, selection works against the recessive allele. This type of selection results in the ***accumulation of the dominant allele***, which, in an extreme situation, might be fixed (p=1).

B) When the recessive allele is favored, the opposite outcome is expected, with a slight difference. Dominant alleles are lost from the population and ***recessive alleles are accumulated***, but the accumulation of the preferred allele is ***slower***. The reason is that in case of full dominance, selection that favors the recessive allele applies to the homozygotes only.

C) When the heterozygote is more fit than either homozygote, it is called ***overdominance*** in fitness. In this situation, selection works against the homozygotes. In such populations, ***equilibrium is reached between the allele frequencies*** according to the fitness of the different genotypes. ***Selection maintains the polymorphism of the population***.

A good example of overdominance is the high frequency of heterozygotes (*Hb-A/Hb-S*) for ***sickle-cell anemia*** in West Africa. Homozygotes for the abnormal allele (*Hb-S*) suffer severely from anemia and die prematurely. Homozygotes for the normal allele (*Hb-A*), on the other hand, frequently die from malaria. Heterozygotes suffer only mildly from anemia, and the presence of the abnormal hemoglobin in their red blood cells protects them against malaria caused by *Plasmodium falciparum*.

D) When heterozygotes are less fit than either homozygote, it is called ***underdominance***. In this case, selection favors an allele when it is common, and the ***population is unstable*** until allele *A* or *a* is fixed. ***Underdominance works against polymorphism***.

A well-known example of underdominance in fitness is ***Rh incompatibility*** in humans. Rh-negative mothers produce antibodies against the red blood cells of their Rh-positive foetus. This is why Rh-positive children born to Rh-negative mothers often suffer hemolytic anemia as newborns. These babies are heterozygotes *(Rh^+/Rh^-).*

Allele frequencies in a population may also reach ***equilibrium when natural selection removes deleterious alleles introduced by repeated mutations***. This balance explains the ***persistence of deleterious alleles*** that cause genetic diseases in the human populations.

Random events

The Hardy-Weinberg equilibrium shows that the amount of genetic variation remains constant from generation to generation in the populations of sexually reproducing organisms. However, this rule applies only to ideal populations.

Real populations consist of a finite number of individuals, and parents may have a small number of offspring. In real populations, even in the absence of disturbing factors, allele frequencies are not reproduced because of sampling errors. In a very small population, random loss of individuals and the alleles they possess may cause even relatively large changes in allele frequencies over one generation. This random change in allele frequencies is called ***genetic drift***.

Sometimes a chance environmental event greatly reduces the population size, and only a small portion of the individuals survive. This causes a ***"bottleneck effect"*** on the population (Figure 139.). The surviving individuals may represent a thoroughly different allele composition than the original population, thus the new population that results from their propagation will have different allele frequencies.

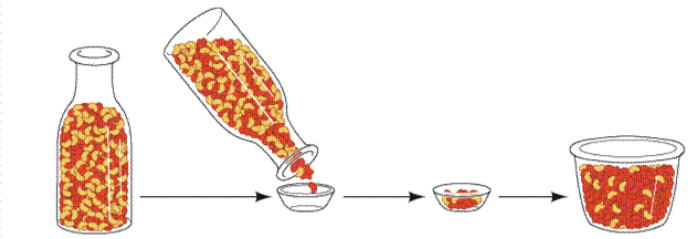


Figure 139. "Bottleneck effect"

A real bottle is indeed suitable for demonstrating the bottleneck effect on a population.

If different populations isolated from an original one undergo divergent changes, the variation present in the original population finally becomes variation between populations.

When a few pioneering individuals break off from a larger population and colonize a new region, the resulting population is unlikely to have all the alleles found in the source population. The resulting change in genetic variation is called the ***founder effect***. The phenomenon is essentially the same as when a large population is reduced by a bottleneck.

An example of the founder effect is the virtually complete lack of blood group B in native Americans. The ancestors of these people broke off from a Northeast-Asian population that had an intermediate frequency of group B. Small groups of individuals crossed the Bering Strait at the end of the last Ice Age, and evolved in isolation from the original population.

8. Quantitative Genetics

8.1 Quantitative traits

Methods of quantitative genetics

Geneticists usually plan their experiments on the base of a ***simple relationship between genotype and phenotype***. In classical genetics, different alleles of a gene produce different, well-distinguishable phenotypes, like e.g. yellow versus green seed color, or red versus white flower. The experimenter can infer the genotype of an individual from its phenotype directly, or at worst, performs some simple crosses to distinguish dominant homozygotes from heterozygotes. This approch is valid, however, only in the world of ***qualitative traits***.

Most variation between organisms is quantitative. Height, weight, shape, color, metabolic activity, reproductive rate are all characteristics that vary more or less continuously over a range. Environmental changes can affect these traits considerably, and several genes are likely to be involved in producing a particular phenotype. ***In case of quantitative traits, the relationship between genotype and phenotype is complex***.

Quantitative genetics applies different methods to study these traits:

Norm of reaction studies are performed to examine the interaction of genotype and environment in the development of a character.

Selection studies can help to determine whether phenotypic differences are based on underlying genotypic differences, or simply the result of environmental changes.

The goal of ***heritability studies*** is to distinguish genetic variance from environmental variance, and to quantify the degree of heritability.

Quantitative trait locus (QTL) studies locate the regions of the genome where loci likely to be involved in determining quantitative traits are situated.

Quantitative genetics applies statistical methods

Quantitative characters do not assort in a simple Mendelian way in crosses, rather they vary continuously. To study quantitative variations, geneticists need to apply statistical methods.

First, a set of quantitative measurements is carried out, and presented as a ***statistical distribution***. For example, the height of people in a group is measured.

Then, the number of individuals observed in each height group is plotted against the measurement classes, producing a ***frequency histogram*** (Figure 140.).

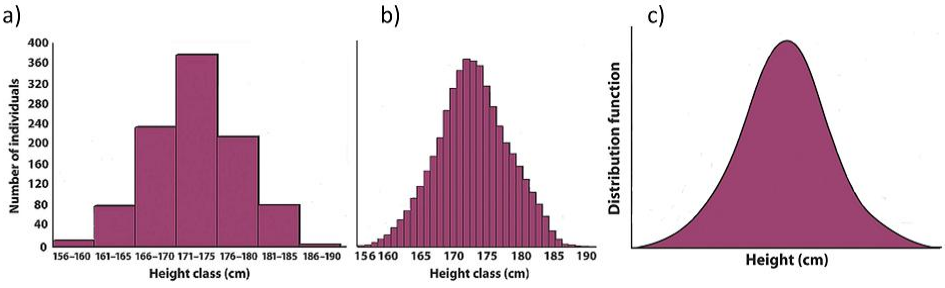


Figure 140. Frequency distribution for height of a group of students

a) Measurement was taken at 5cm class intervals; b) A more accurate measurement with 1cm class intervals.
c) The limiting continuous distribution.

In an ideal situation, unlimited number of individuals can be measured with infinite accuracy. This way the actual frequency distribution of a population could be transformed into a continuous curve (Figure 140c.), which is the ***distribution function*** of the measurement in the population.

The following statistical terms are often used to describe such a distribution:

The "mode" and the "mean" are used to describe the ***central tendency*** of the distribution.

The ***mode*** is the most frequent observation, while the ***mean*** is the arithmetic average of the observation.

How wide or narrow is the variation among the individuals is described by the term "***dispersion***". Dispersion is measured by the ***variance***, which is the average squared deviation of the observation from their mean.

When the ***relation between the measurements*** is addressed, geneticists must calculate their "correlation". ***Correlation*** is the average product of the deviation of one variable from its own mean times the deviation of the other variable from its own mean.

For further explanation of the statistical concepts, please refer to basic textbooks of mathematics.

Quantitative or Mendelian trait?

In classical genetics textbooks, the distinction between qualitative and quantitative traits is usually described by the following criteria:

Qualitative traits appear in a few contrasting phenotypic classes, with a discontinuous variation. The environment affects the phenotypic expression of the genotype only to a negligible extent. The genetic background is easy to explore: usually alleles of a single gene determine the phenotypes, giving Mendelian ratios in crosses.

Quantitative traits, on the other hand, show a continuous variation of phenotypes. Environmental changes often have a profound effect on the phenotypic expression of quantitative traits. Exploring the genetic background is not easy: according to the ***multiple-factor hypothesis***, a large number of genes, each with a small effect, are responsible for producing the different phenotypes.

However, this distinction of the qualitative and quantitative traits is not necessarily satisfying in every situation. Let us elucidate the difference between qualitative and quantitative traits with the help of the following example. A population of plants, containing three different genotypes (*A/A, A/a*, and *a/a*) that affect the height of plants, was grown in an environment where there was some environmental variation.

The three phenotypic distribution of the three plant genotypes are depicted in Figure 141a.

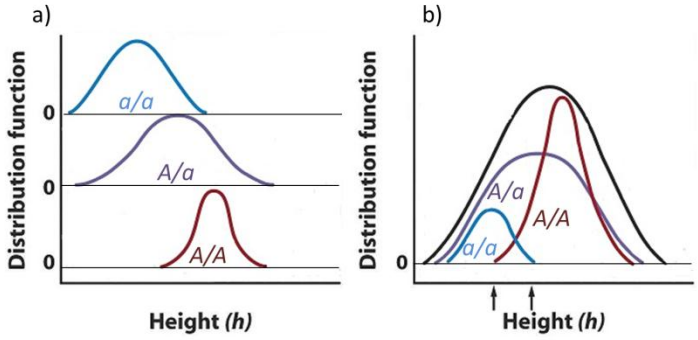


Figure 141. Genotype distribution

a) Phenotypic distribution of three plant genotypes. b) A phenotypic distribution for the total population (black).

Figure 141a shows that there is a difference in mean height, and also in variance between the different genotypes. If we take into consideration that the three genotypes occur in the population in unequal proportions (1*a/a* : 2*A/a* : 3*A/A*), the three genotypic distributions and the phenotypic distribution of the whole population can be presented as in Figure 141b.

Although plant height is determined in this hypothetical population by a single gene, figure 141b reveals that the genotype of individual plants cannot be determined by simple crosses and inspecting progeny ratios. Any plant whose height falls in the range indicated by the arrows can have any of the three genotypes. Because phenotypic distributions overlap so much, a cross between e.g. *A/A* and *a/a* will not give progeny that fall into discrete classes with regular Mendelian ratios but will cover a wide range of phenotypes.

Let us repeat our experiment with this hypothetical plant in a new environment, which is identical for all individuals and increases the growth rate significantly. The results are presented in Figure 142.

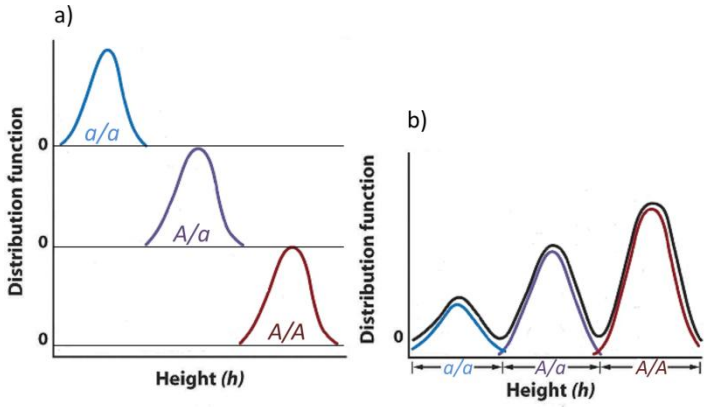


Figure 142. Phenotypic distribution of the same three plant genotypes shown in Figure 141 in controlled environment

The phenotypic variance of each genotype is reduced, because all the plants were grown under the same condition. At the same time, the phenotypic variances between the genotypes are exaggerated due to the more rapid growth. As a consequence, the population is divided into three nonoverlapping phenotypic distributions, to which characteristic genotypes can be assigned. Thus the analysis of this population with Mendelian crosses has become possible.

Making the difference between the means of the genotypes large compared with the variation within the genotypes has converted a “quantitative” character into a “qualitative” character.

The critical difference between Mendelian and quantitative traits is not the number of loci that influence a trait, rather the size of phenotypic differences between genotypes compared with the individual variation within genotypic classes.

8.2 Norm of reaction

Organisms do not live in isolated places but interact with the changing environment during their whole life. Scientists describe the ***relation between genotype, environment and phenotype*** with the term ***norm of reaction***. For a particular genotype, a collection of phenotypes can be gathered that result from the development of that genotype under different environmental conditions. This is called the genotype’s norm of reaction.

Determining norms of reaction

The beginning for any norm of reaction study is having a ***large group of genetically identical test individuals***. Either a homozygous line of individuals must be created, or two different homozygous lines can be crossed, and the resulting heterozygous F_1 offspring analyzed. Both methods result in genetically identical individuals that can be tested in different environments.

To carry out norm of reaction studies is easier with plants than with animals. ***Plants can be propagated vegetatively***, that is, by cutting. The pieces cut from a single plant can be regenerated into whole plants and examined in different environments.

Such a model study was carried out with the yarrow plant, *Achillea millefolium*. After collecting many plants, three cuttings were taken from each individual, and planted at low, medium and high elevation. Then, the heights of plants developed at different elevations were measured (Figure 143.).

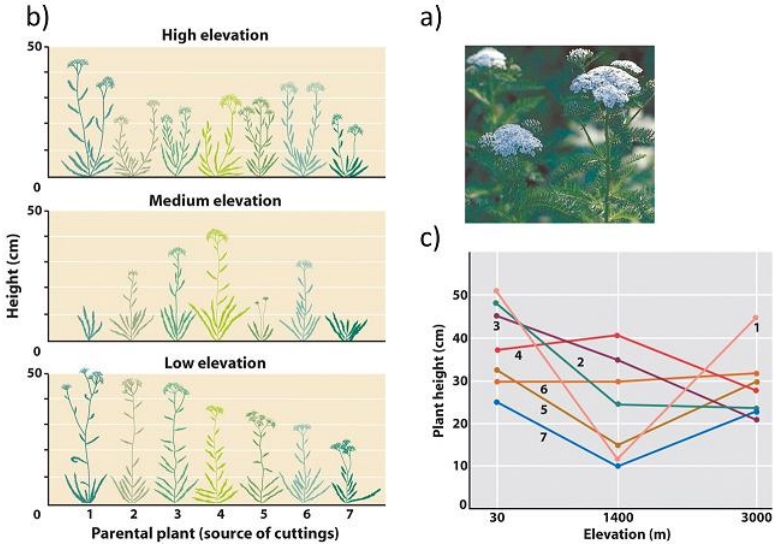


Figure 143. Norm of reaction study with yarrow plant

a) *Achillea millefolium*; b) Norms of reaction to elevation for seven different genotypes; c) Graphic presentation of the norms of reaction

Source: Figure 1-21 and 1-22 in: Introduction to Genetic Analysis, Eighth Edition; © 2005 W. H. Freeman and Company

The results show that there is no such environment (altitude) that is uniformly preferred by all plants. The variation for a single *Achillea* genotype in different environments is so great that the norms of reaction cross each other, and form no consistent pattern.

Producing clones of animals is difficult, and cloning human individuals is completely impossible both from moral and technical aspects. For this reason, we do not know a norm of reaction for any genotype for any human quantitative trait.

For ***domesticated plants and animals***, the technique of ***mating close relatives (inbreeding)*** can be applied. Repeated self-pollination (selfing) in plants, and mating brother and sister in animals can convert a segregating line homozygous. (Segregating lines contain homozygous as well as heterozygous individuals at a particular locus.) However, both methods are slow (Table XI). Selfing reduces the ratio of heterozygotes by 50% from generation to generation, and mating brother and sister reduces heterozygosity even slower.

	Remaining Heterozygosity	
Generation	**Selfing**	**Brother-sister mating**
0	1.000	1.000
1	0.500	0.750
2	0.250	0.625
3	0.125	0.500
4	0.0625	0.406
5	0.03125	0.338
10	0.000977	0.114
20	1.05×10^{-6}	0.014
n	$Het_n = ½\ Het_{n-1}$	$Het_n = ½\ Het_{n-1} + {}^1/_4\ Het_{n-2}$

Table XI. Calculation of remaining heterozygosity in two different mating systems

General experiences from norm of reaction studies

Very few norm of reaction studies have been carried out with natural populations. Researchers usually focus on model systems that are easier to handle, like e.g. *Drosophila*, or they study domesticated species with economic importance, like e.g. corn.

A study mentioned above was carried out with corn, to test grain yield in different environments. Two corn varieties, an older hybrid corn and a new hybrid were compared. When the corn varieties were planted at high density, the new hybrid proved to be superior. High-density-planting was the characteristic farming practice when the new variety was developed. However, comparison of the two varieties at low planting density, which was the characteristic farming practice at the time of developing the old variety, did not show the unequivoval superiority of the new variety. Under the best farming conditions, the old variety performed better. From this, and similar studies the next conclusion can be drawn.

Usually, ***there is no genotype that performs better*** (produces more favorable phenotype) ***than all others under all environmental conditions***. Typically, the differences between genotypes are small, and the directions of the differences change inconsistently with the changing environment. When breeders select a ***"superior" genotype*** in domesticated animals and cultivated plants, this ***is usually a variety that is highly adapted to some specific environmental condition***, and may not show any superiority if the environmental conditions change.

8.3 Heritability

Distinguishing familiality and heritability

Phenotypic variations of different characters can be observed in every natural population. Behind the phenotypic expression of each character there is a gene-mediated developmental process. Genes control the development of every character, but the variation in a character from individual to individual is not necessarily the result of genetic variation. The environment also plays a role. The question is whether genes are involved in the phenotypic differences of individuals or not. A character is called heritable only if there is genetic variation in that character.

If genetic variation influences the phenotypic variation in a trait, biological relatives are expected to be more similar to each other than non-related individuals. Related individuals, however, often live together experiencing the same or very similar environment.

The similarity of relatives may be evidence for genetic variation only if the relatives do not share common environment more than nonrelatives do. Character traits are ***familial if relatives have them in common***, but ***heritable only if the similarity arises from a common genotype.***

In experimental organisms, two general methods can be used to distinguish environmental similarity from genetic similarity (or hereditability). The one called ***marker-gene segregation*** is based on molecular biological experiments. The aim of this method is to show whether genotypes carrying different alleles of a marker gene also differ in their average phenotype for a qualitative character. If the marker genes vary in relation to the character, they are likely to be linked to genes that do influence the character and its variation.

The second, more traditional method is based on carrying out crosses and studying ***phenotypic similarity*** (Figure 144.). In this method, individuals are selected from the two extremes of the phenotypic distribution in the parental generation. Crosses are performed within these two groups and the offspring is raised in common environment. If the phenotypic distribution of the two groups differs significantly, the character difference is heritable.

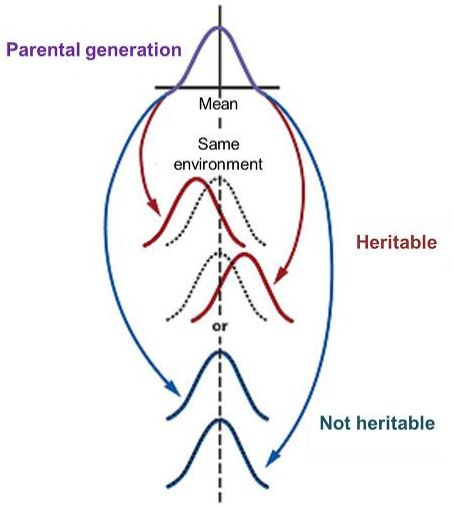


Figure 144. Testing heritability by a traditional method

For explanation, refer to the main text.

Broad heritability

If a character is shown to be heritable, the next step is to quantify its heritability. Any particular phenotype is the result of an underlying genotype and the environment in which this genotype is expressed. Thus, variation between phenotypes in a population comes from two sources. There are genotypic differences, and each genotype exhibits phenotypic variation due to environmental variation. The total ***phenotypic variance*** of the population (s^2_p) is given by the sum of the ***genetic variance*** (s^2_g) and the ***environmental variance*** (s^2_e). (For a more accurate calculation, the possible covariance between genotype and environment (+ 2 cov ge) should also be considered. However, in a planned experiment it can be held at 0 and ignored.)

$$s_p^2 = s_g^2 + s_e^2$$

The measure of heritablity of a character is defined by the part of the total phenotypic varience that arises from genetic variance. H^2, so defined, is the ***broad heritability*** of the character.

$$H^2 = \frac{s_g^2}{s_p^2} = \frac{s_g^2}{s_g^2 + s_e^2}$$

Let us see an example, how heritability in a population can be estimated. After making several homozygous lines, these should be crossed in pairs to create heterozygotes typical of the population. The individuals in these groups have the same genotype, thus measuring the phenotypic variance within each group of heterozygotes will provide an estimate of the environmental variance of the population (s_e^2). This value can be then subtracted from the value of phenotypic variance of the original population (s_p^2) to give the value of genetic variance (s_g^2).
(Though with this method any covariance between genotype and environment in the original population remains hidden, it is sufficient for a rough evaluation of the genetic variance.)

Knowing the values of broad-sense heritability of a character in a population alone is not very useful. If H^2 is not zero, we can conclude that genetic differences have influenced the phenotypic variation among individuals. However, finding zero heritability for a trait still not necessarily means that genes are irrelevant to the trait. It may show that in the particular population and environment studied there is no genetic variation, or different genotypes have the same phenotype.
Determining ***H^2*** can be the first step of a more detailed investigation, but we must keep in mind that this data ***is valid only for the given population and environment***. ***The heritability of a character difference is different in each population and environmental setting***.

A further problem of using H^2 is that this value provides a limited prediction of how much a character can be modified by changing the environment. For the determination of H^2, the phenotypic variance is separated into genetic and environmental components. This distinction, however, does not really separate the genetic and environmental causes of variation, because genotype and environment interact to produce the phenotype.

An erroneous use of heritability is to assume that traits that show high heritability cannot be changed much by environmental changes. This interpretation mixes heritability with unchangeability. A character with high heritability in a population can be subject to great changes resulting from environmental variation.

Narrow heritability

For animal and plant breeders ***heritability in a narrow sense (h^2)*** provides valuable information, because knowing this value they can ***predict whether a program of selective breeding will succeed*** in changing the population. To define h^2, and to understand the importance of the additive component in genetic variance, let us take an example.

In qualitative genetics, the phenotypes of homozygotes and heterozygotes show simple relation generally. In case of full dominance, one unit of the gene product is sufficient to produce full physiological activity which makes the phenotype of homozygous dominants and heterozygotes indistinguishable. The other usual option is that the physiological activity is proportional to the amount of active gene product, creating an exactly intermediate phenotype in the heterozygotes (no dominance).

For many quantitative traits, heterozygotes show partial dominance: their phenotype falls closer to one of the homozygotes. Suppose that alleles *A* and *a* influence the height of an organism. Table XII shows the mean phenotypes and the frequencies of the three genotypes in a hypothetical population.

	a/a	*A/a*	*A/A*
Phenotype	10	18	20
Frequency	0.36	0.48	0.16

Table XII. Parameters of a hypothetical population
For explanation, refer to the main text.

The phenotypic means of the three genotypic classes are different, so there is genetic variance in this population. Let us calculate the average effect of allele *A* and *a* on the phenotype. To do this, the occurrence of the allele in question must be counted, and multiplied by the heights of the individuals in which this allele appears.
Thus, the average effect of the two alleles can be calculated as follows:

$$\bar{a} = \frac{2\,(0.36)(10) + 1\,(0.48)(18)}{2\,(0.36) + 1\,(0.48)} = 13.2 \text{ cm}$$

and

$$\bar{A} = \frac{2\,(0.16)(20) + 1\,(0.48)(18)}{2\,(0.16) + 1\,(0.48)} = 18.8 \text{ cm}$$

In this population, the average difference in effect between the two alleles is 5.6 cm. This is the additive effect that accounts for some of the variance in phenotype. In addition, there is some dominance.

Generally, the total genetic variance in a population can be subdivided into ***additive genetic variance*** (s_a^2) and ***dominance variance*** (s_d^2). The former arises from the average difference between the carriers of *a* alleles and the carriers of *A* alleles, the latter results from the fact that heterozygotes are not exactly intermediate between the homozygotes.

$$s_g^2 = s_a^2 + s_d^2$$

Thus the total phenotypic variance can be defined as the sum of additive genetic variance, dominance variance, and environmental variance:

$$s_p^2 = s_g^2 + s_e^2 = s_a^2 + s_d^2 + s_e^2$$

With the introduction of s_a^2, we can define ***heritability in the narrow sense (h^2):***

$$h^2 = \frac{s_a^2}{s_p^2} = \frac{s_a^2}{s_a^2 + s_d^2 + s_e^2}$$

The effect of selection depends on the amount of additive genetic variance. The greater h^2 value indicates that greater fraction of the difference between the selected parents and the population as a whole will be preserved in the next generation. ***Breeders must rely on h^2 and not H^2 when planning a selection program.***

Estimating narrow heritability

Estimating narrow heritability (h^2) of a trait can be done by calculating the regression line of offspring measurements (y) on ***midparent values*** (y) (Figure 145.). In practice, we must plot the phenotype of offspring against the average phenotype of their two parents. (This latter is the midparent value.) The slope of the regression line relating offspring value to the midparent value provides an estimate of h^2.

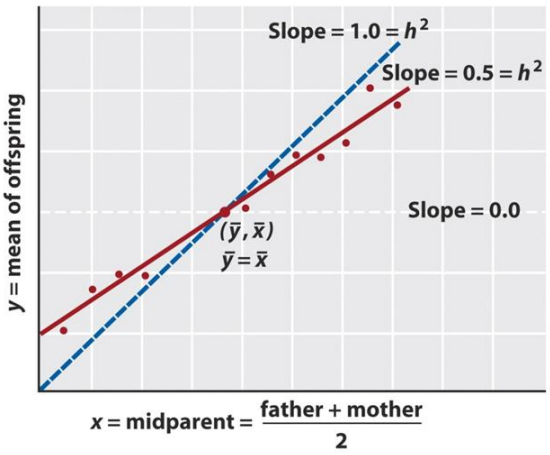


Figure 145. Estimating narrow heritability

Slope = 1 = h^2: The phenotype is additively inhereted with complete fidelity; the values measured in the offspring are identical with the midparent values.
Slope = 0: The offspring has no heritable similarity to their parents; all parents have offspring showing the same phenotypic value.
0 < slope > 1: In practice, heritability is not perfect. For instance, very tall parents tend to have somewhat shorter, while very short parents tend to have somewhat taller children than they themselves are.

Because the slope of the regression line is a correct estimate of h^2, it is possible to use it in the prediction of the effect of artificial selection:

$$\text{selection response} = h^2 \times \text{selection differential}$$

The ***selection response*** is the difference between the offspring of the selected parents and the mean of the parental generation, while the ***selection differential*** is the difference between the selected parents and the mean of the entire population of parents.

Considering that

$$h^2 = \frac{\text{selection response}}{\text{selection differential}}$$

carrying out selective breeding for one generation, and comparing the selection response with the selection differential, provides us with an alternative way to estimate h^2.

Artificial selection

Artificial selection is a very efficient way to change the phenotypic composition of a population. Man has been experimenting with it for centuries in order to "improve the quality" of economically important plants and animals. With the advance of genetics, this activity is done with more and more awareness.

The usual method of selection for a trait showing continuous variation is ***truncation selection***. The individuals in a given generation are pooled, the phenotypic values are measured, and only those individuals are chosen as parents for the next generation which are above (or below) a given phenotypic value. This value is called the truncation point.

Artificial selection programs usually encounter the same obstacle. As the population becomes more and more extreme, its viability and fertility decreases. If this loss of fitness is another phenotypic effect of those genes that influence the character we want to improve, the program cannot be continued after a certain point. However, if the loss of fitness is due to sterility genes that are linked to the genes under selection, there might be a solution. Breeders can suspend selection for a few generations, and hope that recombination by chance frees the genes under selection from the accompanying sterility genes.

Using heritability values in breeding

Although h^2 always applies to a particular population and environment, breeders still can find it useful, because in practice they work with a particular group of animal in a given environment. If h^2 is high, there is a high parent-offspring correlation, and they can expect that a high portion of the superiority of the selected parent will appear in the offspring.

If h^2 is low, breeders consider h^2 as well as H^2. When both h^2 and H^2 are low, there is a large proportion of environmental variance compared with genetic variance. In this case, breeders use ***family selection***. Breeding pairs are allowed to produce several trial progeny, and parental pairs are selected for further breeding on the basis of the average performance of those progeny.

If h^2 is low but H^2 is high, there is not much environmental variance. In this situation, the additive genetic variance is small compared with the dominance variance. Breeders use the ***hybrid-inbred method***, which was used widely for improving corn in North America. First, many inbred lines are created by selfing. Then, these inbred lines are crossed in many different combinations. From the resulting hybrid lines the best one is selected again, and subjected to inbreeding. Repeating this cycle several times breeders select for additive effect as well as for dominance effect.

Quantitative trait loci

Locating the genes that influence quantitative traits in the genome is a complicated process whose understanding requires advanced knowledge in molecular biology. For this reason, the discussion of quantitative trait loci will be restricted to some general remarks.

It is not possible to identify all the genes that influence a given character by using genetic methods only, because genetic analysis can detect a gene only if there are allelic variations. Molecular analysis, in contrast, can identify genes even in the absence of variation, provided the gene product can be identified.

Sometimes prior knowledge about the biochemistry or development of an organism allows focusing on a locus that is likely to be involved in the development of a phenotypic variation. This locus is called ***candidate gene***.

The actual genes segregating for a quantitative trait cannot be identified in most cases. Usually these genes, the ***so-called quantitative trait loci***, can be localized to certain regions of the genome only. This localization is done with the help of ***marker genes*** that produce well-detectable phenotypes and are linked to the quantitative trait loci.

9. Agricultural Biotechnology

9.1 The concept of biotechnology

What is biotechnology?

The term "biotechnology" was first used by a Hungarian agricultural economist, Károly Ereky, in 1919. What he meant by this term is the production of products from raw materials with the aid of living organisms. Since then, several other definitions have been used to describe biotechnology.

According to the definition given by the European Federation of Biotechnology, ***biotechnology is "the integration of natural sciences and organisms, cells, parts thereof, and molecular analogues for products and services"***. Whatever definition we use, the following characteristics are generally included in the concept of biotechnology:

The ***application of living systems*** (microbial, animal or plant cells) ***or materials that are produced by or produced from living organisms*** (e.g. enzymes, biomass) is always a key element in the definition. These cells or their products are used to synthesize, breakdown or transform material, so that the ***final product should meet some practical human need***.

The huge advance achieved in molecular biology in the second half of the twentieth century and understanding the fundamentals of life processes have ***greatly contributed to the development of modern biotechnology***. However, biotechnology is not a single biological discipline rather an ***interdisciplinary activity***. Biotechnology can draw upon a wide array of relavant fields, such as microbiology, biochemistry, molecular biology, genetics, cell biology, immunology, etc, and many aspects of biotechnology have arisen through the ***interaction between*** various parts of ***biology and engineering***.

Modern biotechnology is at the ***interface between academic sciences and industry***. On the one hand, it is based on basic sciences; on the other hand, it can inspire the development of sciences and provide them with new methods. An important distinction between biology as an academic science and biotechnology is their scale of operation. Biotechnology has to ***scale-up biological processes for industrial use***.

Much of biotechnology revolves around microbial processes, and so, many biotechnological processes can be broken down into three main parts. During the first part researchers deal with ***obtaining the best "biological catalyst"*** (e.g. best microbial strain, proper cell culture, etc.) for a specific function or process. The aim of the second part is ***creating the best possible environment*** for the catalyst to perform. The third part, also called downstream processing, is concerned with the ***separation and purification of the product or products.***

Brief history of biotechnology

The historical development of biotechnology can be divided into two parts: the era of traditional biotechnology and the era of modern biotechnology.

Traditional biotechnology dates back to prehistoric times: humans used microbes to produce food and beverages, and modified plants and animals through progressive selection for desired traits for thousands of years. Traditional biotechnology was based on ***observation and practical experience***. Yeasts were used to produce beer, wine and bread before even knowing them about anything (Figure 146.).

Fermentation technologies, which comprise the main part of traditional biotechnology, gained a further momentum for development when Pasteur (Figure 147.) demonstrated the fermentative abilities of microorganisms.

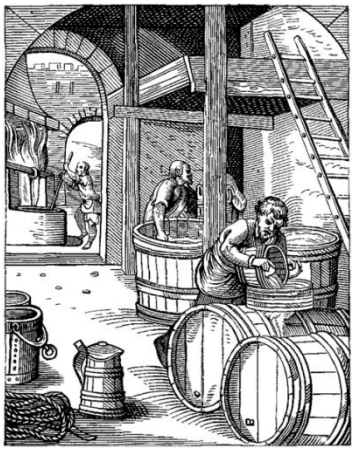


Figure 146. The brewer, a 16th century engraving

https://commons.wikimedia.org/wiki/File:The_Brewer_designed_and_engraved_in_the_Sixteenth._Century_by_J._Amman.png
Jost Amman, Public domain, via Wikimedia Commons

Figure 147. Louis Pasteur (1822-1895)

Pasteur was a French chemist and microbiologist who discovered the principles of vaccination, microbial fermentation and pasteurization. He is honored as "the father of bacteriology and microbiology".
https://en.wikipedia.org/wiki/File:Louis_Pasteur,_foto_av_Paul_Nadar,_Crisco_edit.jpg

Biotechnological processes initially developed under non-sterile conditions. By the end of the nineteenth century, several products, like e.g. ethanol, acetic acid, butanol and acetone were produced by open microbial fermentation processes. Fermentation capacities were applied for waste-water treatment and municipal composting of solid wastes.

Further advance in biotechnological processes came with the introduction of sterile conditions in the cultivation of microorganisms. In the 1940s, this progress allowed the mass production of antibiotics, amino acids, organic acids, enzymes, etc., but microbial strains were still improved with traditional technologies.

The revolution of ***modern biotechnology*** began in the 1970s and early 1980s when scientists began to acquire the tools for the ***alteration of the genetic constitution of living organisms***. Recombinant DNA technologies, together with protoplast fusion, have allowed the programming of the biological properties of organisms. By nowadays it has become obvious that ***biotechnology is a leading field of technological development in the twenty-first century***.

Main branches of biotechnology

Biotechnology permeates many areas of modern life (Figure 148.). It has many applications in ***medicine***. Modern biotechnology can be used to produce ***pharmaceutical drugs*** relatively easily and cheaply. The first example is insulin, which is widely used for treatment of diabetes, and was extracted from slaughtered animals before inventing its production by recombinant DNA technologies. Biotechnology has been contributing to the production of several traditional and new drugs for decades.

To further expand the use of biotechnology in medicine, a new research field, ***pharmacogenomics*** has arisen recently from the combination of pharmacology and genomics. The aim of pharmacogenomics is to optimize drug therapy to the patients' individual needs.

Another important field in medicine is ***genetic testing***. Genetic tests are usually used to find changes that are associated with inhereted disorders. Testing can help to determine the chance of developing or passing on a genetic disorder.

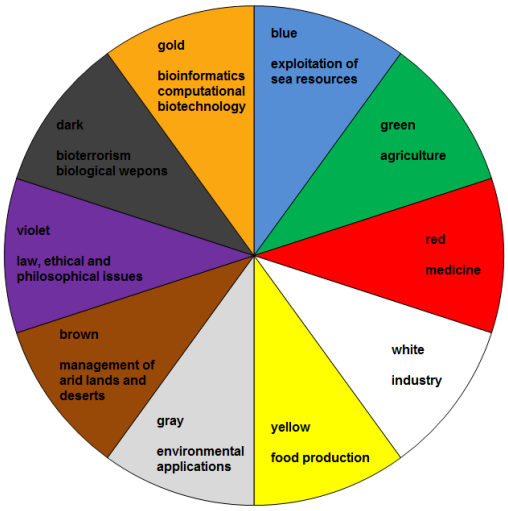


Figure 148. "The colors of biotechnology"

A "color code" was coined to idenfy the different branches of biotechnology.

In ***agriculture***, a major field of application of modern biotechnology is the production of genetically modified organisms, especially ***genetically modified crops***. (We will return to this topic in more detail later.) Further examples of agricultural application of modern biotechnology include the ***production of pharmaceutical agents and biofuels***, as well as the use of biological systems in ***bioremediation***.

Biotechnology has a wide-spread application in ***industry,*** including the most traditional application in ***industrial fermentation***. Industrial biotechnology often uses microorganisms and enzymes to produce the most different types of products, such as chemicals, food, detergents, biofuels, etc. ***Synthetic biology*** is a new and essential cornerstone of industrial biotechnology. Synthetic biology is used to engineer model organisms such as *E. coli* or *Saccharomyces cerevisiae* to better suit the challenges of modern industrial biotechnology.

Environmental applications of biotechnology aim to reduce environmental waste and to provide environmentally safe processes such as biofiltration and biodegradation.

Plant biotechnology

After specifying biotechnology in a broad sense, let us focus our attention to a field in agricultural biotechnology: plant biotechnology. Plant biotechnology utilizes the results of several disciplines, such as cell biology, histology, tissue culture, plant physiology, genetics, molecular biology and immunology. ***Modern plant biotechnology can be defined as the use of tissue culture and genetic engineering techniques to produce genetically modified plants that exhibit new or improved desirable characteristics.***

Based on this definition, two major branches of plant biotechnology can be described. Cell and ***tissue culture***, also called *in vitro* techniques, deal with the propagation and regenaration of cells, tissues, organs and embryos under sterile, controlled conditions on nutrient culture medium of known composition. ***Genetic engineering*** creates valuable plants with new traits through genetic manipulation of the hereditary material.

Plant cells are at the center of plant biotechnology, because they are ***totipotent***: from a single genetically modified cell a whole plant can be regenerated.

9.2 Plant tissue culture

Some basic terms in plant tissue culture

Plant tissue culture, as defined before, is a quite general term, as it can refer to *in vitro* propagation of all types of plant cultures.The first decision a researcher or breeder has to make is choosing the explant.

The "explant" is the excised part of the intact plant organism. Explants can be taken from many different parts of the plant; portions of shoots, leaves, stems, flowers, roots, as well as single undifferentiated or mature cells can be used to initiate a plant culture. Special care should be taken that the plant selected as explant source be ***healthy and actively growing***. Insec- and disease-free greenhouse plants are preferred starting material for tissue culture. Also working with meristems has an advantage, which is the high probability of excluding plant pathogens, especially viruses that can be present in the donor plant.

Aseptic technique is very important when working with tissue cultures. Sterile tissues are used, or plant pieces are subjected to surface disinfection to remove contaminants, taking care that minimal damage to the plant cells occur. Throughout the whole work, using sterile instruments, culture media and containers, as well as applying standard bacteriological transfer procedures are essential. Plants tissues are usually prepared for tissue culture under air-filtered laminar flow cabinets also called tissue culture hoods (Figure 149.).

Figure 149. One of the many different types of laminar flow cabinets

https://commons.wikimedia.org/wiki/File:Laminar_flow_cabinet_Microbiology_Department.jpg
Chainwit., CC BY-SA 4.0 <https://creativecommons.org/licenses/by-sa/4.0>, via Wikimedia Commons

Tissue cultures are grown in ***culture media*** that can be used in solid and liquid form. Culture media generally contain ***major salts, minor salts, vitamins, sucrose and hormones***. The exact composition is adjusted to the physiological needs of the cultivated plant material and the goal of regeneration.

Plant hormones are especially important for the normal development of plant tissues. The main plant hormones used in tissue culture are the ***cytokinins and auxins*** (Figure 150.). Cytokinins stimulate DNA synthesis and cell division, while auxins stimulate cell expansion, particularly cell elongation. Plant hormones function in relation to each other, thus hormone balance is more important than absolute concentration of any one hormone. In actively dividing tissues both cell division and cell expansion occur. The amounts of plant hormones in the culture medium have a profound effect on the development of the tissues that grow from the initial explants. An excess of auxins tend to promote root proliferation, while an excess of cytokinins tend to promote shoot development.

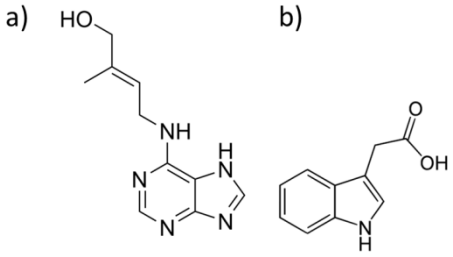


Figure 150. Plant hormones

a) Zeatin, an adenine-type cytokinin named after the genus of corn, *Zea*;
b) Indole-3-acetic acid, the most abundant auxin in plants.

In tissue culture, not only the exact composition of the culture media, but also the ***culture conditions such as temperature, light, and humidity are defined.***

The most commonly used tissue explants are the meristematic ends of the plants like the ***stem tip, axillary bud tip and root tip*** (Figure 151.). The advantage of using meristematic cells is that these cells are undifferentiated or incompletely differentiated. They are ***totipotent*** and capable of continued cell division. Starting with meristematic cells and using proper culture media and culture conditions, tissue differentiation and organ development can be directly induced.

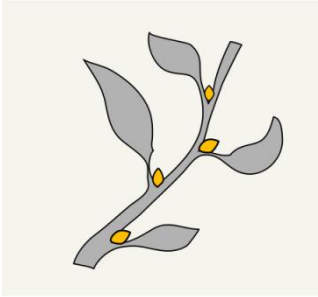


Figure 151. Location of axillary buds

Sources of meristematic cells, frequently used in micropropagation of plants.
https://commons.wikimedia.org/wiki/File:Plant_morphology_buds_axillary.png
Mariana Ruiz Villarreal LadyofHats, CC0, via Wikimedia Commons

Plant tissue cultures can be established not only from meristematic cells, but also from several parts of the intact plant. However, using differentiated tissues raises a problem. Differentiated cells/tissues first must be properly treated to undergo a process called ***dedifferentiation***. Dedifferentiated cells become less specialized; on solid medium they form an ***unorganized tissue mass called callus*** (Figure 152.). Tissue culture media designed to produce callus usually contain elevated levels of cytokinins and auxins.

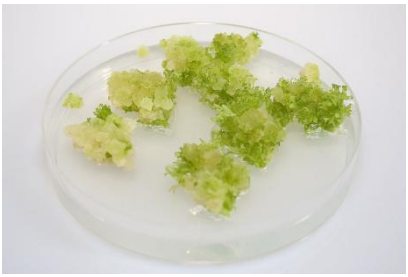


Figure 152. Callus of tobacco (*Nicotiana tabacum*)

https://commons.wikimedia.org/wiki/File:Callus1.jpg
Igge, CC BY-SA 3.0 <https://creativecommons.org/licenses/by-sa/3.0>, via Wikimedia Commons

Callus formation is central to many investigative and applied tissue culture procedures. Callus can be multiplied, and after inducing differentiation and regeneration, used to ***clone numerous whole plants***. Different genetic engineering protocols employ callus initiation after the designed DNA construct has been inserted into the plant cell. ***Transgenic plants are then regenerated from the transformed callus.*** Callus can be used to form ***suspension cultures***. These cultures contain individual plant cells and small cell clusters (microcalli) that have detached from the original callus inoculums due to the intensive agitation of the liquid medium. Suspension cultures can be used as ***inoculums for plant bioreactors***, in which valuable plant products are produced.

Applications of plant tissue cultures

Plant tissue cultures are used in basic research as well as in many applied aspects in plant science. Without being exhaustive, given is below a list of potential applications.
Plant tissue cultures are used

- To investigate basic scientific questions such as the totipotency of plant cells or the roles of hormones in cytodifferentiation and organogenesis
- To create transgenic plants for scientific investigation as well as for commercial distribution
- To investigate the molecular biology and gene regulation of genetically-engineered plant cells
- To make suspension cultures for bioreactors, in which valuable compounds like plant-derived secondary-metabolites and recombinant proteins can be produced
- To produce large numbers of identical plants (clones) for commercial distribution in much shorter time than it is possible by traditional methods
- To cross distantly related species either by protoplast fusion and the regeneration of the novel hybrid, or by rescuing the embryo resulted from cross-pollination
- To produce clean (virus-free) plant material from virused stock for propagation and conservation
- To large-scale production of artificial seeds through somatic embryogenesis
- To conserve rare or endangered plant species

In vitro micropropagation

Basic methods for plant micropropagation were developed in the late 1950s and 1960s. Since then, micropropagation has become a ***widely-used practice for the rapid multiplication of plant stock material*** for commercial purposes. The most frequently used micropropagation method utilizes axillary shoot proliferation. The procedure can be divided into four stages.

1. The first stage is the ***establishment of axenic cultures***. This starts with selection of the mother plant, and surface sterilization of the explant material. Axillary buds and meristems are preferred explant materials. The sterilized plant tissue is placed on a growth medium in which relatively high concentration of cytokinins promotes shoot formation from the axillary buds.

2. The second stage involves the ***multiplication of plant material***. Shoot proliferation and multiple shoot production are promoted with proper modification of the culture medium, e.g. hormones are often added that cause the plantlets to produce many small offshoots. These are then separated and transferred to new culture medium in order to maximize the quantity of shoots produced.

3. During the third stage ***plantlets produce roots and undergo "hardening"*** (Figure 153.). Rooting media usually contain a high auxin/cytokinin ratio. "Hardening" means a kind of preparation of the regenerated plants for transplanting from *in vitro* (controlled test-tube environment) to *ex vitro* (glasshouse, then natural growth environment) conditions. The high humidity, low light and warm temperature experienced in "*in vitro*" cultures are ideal for rapid tissue propagation but very different from the natural conditions, to which the regenerated plants have to adopt gradually. For example, they have to develop cuticula on the leaves and "learn" to use their stomata.

Figure 153. Plant tissue cultures, *in vitro* conditions

https://commons.wikimedia.org/wiki/File:Plant_tissue_cultures,_National_Center_for_Genetic_Resources_Preservation,_USDA.jpg
USDA, Lance Cheung, Public domain, via Wikimedia Commons

4. The final stage of micropropagation is called ***acclimatization***, and involves the ***transfer of regenerated plants to soil under natural conditions***. Direct transplantation of plants from *in vitro* conditions into soil results in a low survival rate, indicating the importance of gradual "hardening".

Haploid cultures

In plant breeding, tissue culture methods can be applied to produce haploid plants. Haploid plants offer important advantages for plant breeders. Haploid plants are smaller than diploids that can be an advantage when propagation in large volumes is needed. What is more important, the haploid state allows ***easy identification and isolation of mutants***.

Haploid plants are valuable tools in programs designed for genetic modification of plants. After transformation of haploid cells, ***transgenic plants can be produced*** in one step

by duplication of the chromosome set. ***Doubled haploids*** produced this way are homozygous. By conventional breeding practices, homozygosity is achieved via inbreeding that has to be repeated for several generations. Doubled haploid production reduces the time and cost of cultivar development, thus greatly improves the economics of plant breeding.

Then selected homozygous male and female plants can be chosen for the production of the desired ***F_1 hybrid plants***. These plants often exhibit so-called ***hybrid vigor***, also called ***heterosis***, which is manifested in much higher yield compared to either parent.

Methods for haploid production: *In vitro* production of haploid plants can be classified into several categories.

1. The most common and most important method for haploid production is the creation of ***androgenetic haploids***. It is known since the 1960s that immature pollen grains (microspores), either in the form as isolated cells or confined within the anther wall, can develop into haploid embryos (Figure 154.). In another word, ***in androgenesis immature pollen grains are induced to switch to the sporophytic mode of development***. Two methods exist for in vitro production of androgenetic haploids: the ***anther and the pollen culture***.

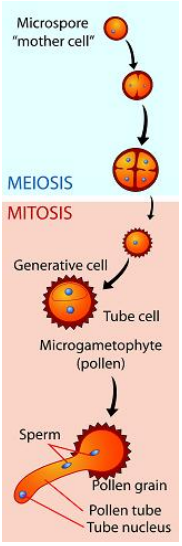


Figure 154. Pollen development

Based on: https://commons.wikimedia.org/wiki/File:Pollen_Tube.svg
SiliconProphet, CC BY-SA 4.0 <https://creativecommons.org/licenses/by-sa/4.0>, via Wikimedia Commons

Preparing an ***anther culture*** is relatively simple and efficient technique. Flower buds are surface sterilized; the anthers are excised from the buds under aseptic conditions, and are cultured on semi-solid or in liquid medium. The cultures are exposed to pretreatments and kept under appropriate conditions. Depending on the plant species and the culture conditions, the androgenic pollen either develops directly into embryo or forms callus tissue. After pollen embryos or calli become visible, the cultures are placed under appropriate conditions for organogenic differentiation.

The other method is the mechanical isolation of pollen grains. Proper composition of culture medium, pretreatment of pollen grains, and proper plating density are the critical factors for the induction of androgenesis. Though the nutritional requirements of isolated pollen are more complex than that of cultured anthers, pollen cultures have advantages over anthers cultures:

a) Isolated pollen can be genetically modified before culturing and the desired genotype can be selected at an early stage of development.
b) Androgenesis is generally more efficient in pollen cultures than in anther cultures.
c) Pollen cultures provide an excellent system for studying the switch from gametophytic to sporophytic development.

2. A possible alternative method for production of haploid plants is the ***in vitro induction of gynogenesis***. This can be achieved by preparation of ovules before pollination, and the induction of gynogenetic development in the culture.

3. The third possible way of generation of haploid plants utilizes the ***selective chromosome elimination that follows certain interspecific hybridization.*** The phenomenon was discovered in barley, when crosses between *Hordeum vulgare* and *H. bulbosum* resulted in selective elimination of the chromosomes of *H. bulbosum* parent. This method is routinely used in wheat and other cereal breeding programs. Haploids of the desired species are produced after pollination with maize pollen. The method involves a phase of embryo rescue in vitro; finally, diploid plants are obtained after colchicine treatment.

Diploidization: Haploid plants are sterile, because in haploids normal meiosis cannot occur. To obtain gamete-producing fertile diploids, the ***chromosome set of haploid plants must be doubled.***
In some plants, especially in those ones where androgenesis occurs via callus development, spontaneous chromosome duplication occurs at a high rate. In such cases, only the identification and selection of diploid plants is needed. (The ploidy level of cells or embryos can be determined by simple chromosome staining.)
When the level of spontaneous diploidization is low, different methods are used to diploidize the haploid plants. The essential step of each procedure is ***treatment with colchicine***. Colchicine is a drug, isolated from the autumn crocus (*Colchicum autumnale*) originally, which induces diploidization by the prevention of microtubule polymerization during mitosis.

9.3 GMOs in agriculture

GMO

GMO means: "genetically modified organism". ***A genetically modified organism is any organism whose hereditary material has been altered by the method of genetic engineering.*** Sometimes the expression of ***"transgenic organism"*** is used as an alternative name. The term transgenic organism emphasizes that a GMO contains a gene which was isolated originally from another organism.

Bacteria are much easier to handle in genetic manipulations than plants or animals, so they represent the first GMOs. The application of ***transgenic microorganisms*** is widespread. Bacteria and yeasts are frequently used in heterologous protein expression, mainly for the production of pharmaceutical drugs. Transgenic bacteria are also used for bioremediation, in environmental biotechnology, and for several other purposes.

Transgenic crops are the product of modern plant breeding. The first GMO crops were used as animal and human food.

Transgenic animals are not only of economic importance, but also play an important role in basic research.

Introducing foreign DNA into plant cells

The oldest branch of biotechnology, the cultivation of plants and husbanding of animals, was based on selective breeding (artificial selection) of these organisms for several centuries. Until recently, breeders had to select varieties with desired phenotypes that had existed in nature through mutational variation, and tried to introduce the desirable genes into a widely used variety of that crop through deliberate crosses. However, this method, partly, is a question of good luck: behind the favorable character there may be complex genetics, the outcome of crosses is hard to predict, and to maintain a good combination as a pure-breeding line is difficult. Moreover, traditional crop plant breeding takes a long time.

Recombinant DNA technology offers ***advantages*** over the traditional breeding methods: It allows breeders to ***choose specific genes***, which makes the process more precise, and helps to avoid the incorporation of unforeseen genes. Theoretically, ***any gene from any organism can be used***, thus the possibilities to develop new characteristics is almost limitless. Researchers can work with cells in the laboratory, and then regenerate a whole plant, which makes plant breeding ***much faster***.

Nowadays it is possible to generate transgenic plants from every major crop plant species. There are several methods for introduction of an altered gene into an organism; for plants, two approaches have survived the test of time: ***transformation using Agrobacterium*** as a vector, and ***particle bombardment***.

Agrobacterium-mediated transformation

Agrobacterium tumefaciens is a Gram-negative ***soil bacterium that infects dicotyledonous plant species***. This bacterium causes ***crown gall disease***, in which the infected plant produces uncontrolled growth, normally at the base of the stem (Figure 155.). The key to tumor production is a large, circular DNA plasmid, the ***Ti (tumor-inducing) plasmid***. Upon infection, a part of the Ti plasmid called ***T-DNA (transfer DNA)*** is transferred and ***inserted***, more or less random, ***into the genome of the host plant.***

Figure 155. Crown gall on a *Kalanchoe* plant

The natural behavior of the Ti plasmid makes it well suited to the role of plant vector. But to re-design it as a vector for plant transformation, first scientists had to reveal the structure of the Ti plasmid (Figure 156.), and understand its function in the process of infection.

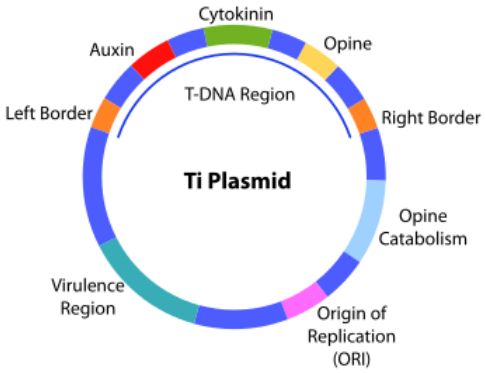


Figure 156. Simplified representation of the major regions of the Ti plasmid

https://commons.wikimedia.org/wiki/File:Ti_plasmid.svg
Mouagip This W3C-unspecified vector image was created with Adobe Illustrator., CC0, via Wikimedia Commons

Agrobacterium predominantly infects plants at a wound site. The damaged plant cells synthesize and secrete unusual phenolic compounds such as acetosyringone, which activate a set of ***virulence genes*** (*vir* region) on the Ti plasmid. The products of *vir* genes are necessary for attaching the bacteria to the plant cell, for the production of a ***single-stranded copy of the T-DNA***, for the transfer of the T-DNA into the plant cell, and for its ***integration into the plant genome***. Integration of the T-DNA leads to plant expression of bacterial genes that encode proteins for the synthesis of opines and plant hormones. The ***opines*** synthesized by the plant tumor cells constitute a food source for the bacterium that neither the plant cells nor any other soil-dwelling microorganism can metabolize. This creates a biological niche for *Agrobacterium* within the tumor tissue. Overproduction of ***plant hormones*** (cytokinins and auxins) ***promotes tumor formation*** in the infected plant cells.

Natural Ti plasmids are usually too large to be easily manipulated. Properly engineered Ti plasmids are created in steps. Several intermediate cloning steps take place in *E. coli*. Tumor inducing genes are removed from the T-DNA, thus creating a ***"disarmed" Ti plasmid*** that is unable to induce a tumor. The transgene is inserted between the border regions of the T-DNA as an ***expression cassette***. In addition to the protein coding region, this cassette contains a plant promoter and a transcriptional termination signal. The transferred region contains another cassette as well, which serves for the selection of transformed plant cells. Usually, this is an ***antibiotic resistance marker***. The *vir* genes that are necessary for the activation of the T-DNA are either integrated into the chromosome of *Agrobacterium*, or they are placed on a helper plasmid.

In the next step, the *Agrobacterium* strain carrying the properly engineered Ti plasmid is used to infect plant tissues. Tobacco is frequently used as a test plant. Punched-out leaf discs are immersed in a bacterial suspension, and then the leaves with the bacteria are transferred onto a proper medium to allow the infection to happen. In infected cells, the T-DNA carrying the transgene and the marker gene is inserted into a plant chromosome. If the leaf discs are placed on a medium that contains antibiotics, only transformed cells that are resistant due to the integrated antibiotic resistance marker can divide. These cells grow into a clump or callus that can be induced to form shoots and roots. The small plantlets are then transferred to soil, where they can develop into transgenic plants (Figure 157.).

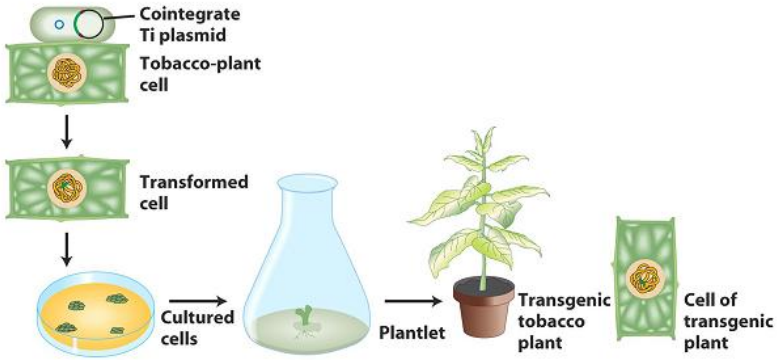


Figure 157. The generation of transgenic plant

Source: Figure 11-30 in Introduction to Genetic Analysis, Eighth Edition; © 2005 W. H. Freeman and Company

Integration and proper functioning of the transgene are controlled at four levels: (1) Experiments verify the integration of the transgene into the genome. (2) Gene expression is monitored at the RNA level and (3) the protein product is detected. (4) When the transgene confers a detectable phenotype, this is also inspected.

The first genetically engineered plant was tobacco, reported in 1983. Since then, many GM plants have been created using *Agrobacterium*-mediated transformation. Initially, this method proved to be less successful for monocotyledonous plants. For the transformation of monocots like maize or wheat, scientists used gene gun, also called particle bombardment.

Gene gun or particle bombardment

A ***gene gun*** is a device used to deliver transgenes to cells (Figure 158.). To get DNA into the cell, heavy ***metal particles are coated*** with the gene of interest. Tungsten particles were used first, but gold or silver particles are also suitable. Gold, being less cytotoxic, is actually favored over tungsten.

The coated particles are placed on a support and accelerated to high speed using a pulse of high pressure helium. Particles hit the target tissues that are placed in an evacuated chamber; this is why the method also called ***particle bombardment***.

Figure 158. A gene gun

This device is called a Helios PDS 1000/He Biolistic Particle Delivery System, sold by the BioRad.
https://commons.wikimedia.org/wiki/File:Genegun.jpg

Particles can penetrate several layers of cells. The target is often a callus of undifferentiated plant cells. When the DNA is released from the particles in the surviving cells, the gene can be transiently expressed, or more rarely, it becomes integrated into the nuclear genome. If the delivered DNA construct contains a selectable marker, stably transformed cells can be selected and cultured using tissue culture methods. Finally, transgenic plants can be regenerated.

Gene guns are most widely used in plants, especially for the generation of ***transgenic cereals***, though the method can be applied to transform human and animal cells as well.

Application of transgenic technologies in plants

Since the production of the first transgenic plant, the application of transgenic technologies has become common in plant breeding. Nowadays, economically important GM crops can be classified into three generations.

1. The ***first generation*** of GM plants carries transgenic constructs that confer ***better production traits***, e.g. herbicide resistance, insect/disease resistance.

2. The ***second generation*** of GM plants carries ***stacked genes for multiple traits***. These involve engineering of the ***metabolic pathway*** and modification of the ***development*** of the plant.

3. The ***third generation*** of GM plants is tailored for ***specific end users***; plants are intended to be used as ***bioreactors*** producing important molecules for the pharmaceutical, feed and plastics industries.

Let us review some of the most remarkable examples.

Herbicide resistance

Weeds compete with crops for space, light, water and nutrient, and they can act as a reservoir for crop pathogens. Thus, it is essential to take efficient action against weeds to achieve abundant crop yields. In most large-scale agricultural systems, traditionally herbicides are used to control weeds.

"Roundup" is a widely used and very efficient ***weed killer*** (Figure 159.). It kills dozens of the most prevalent weeds, but also kills the crop plants, so great care must be taken with its use. A reasonable solution is to make crops herbicide resistant, so the herbicide could be applied to the fields at any time without harming the crop plants.

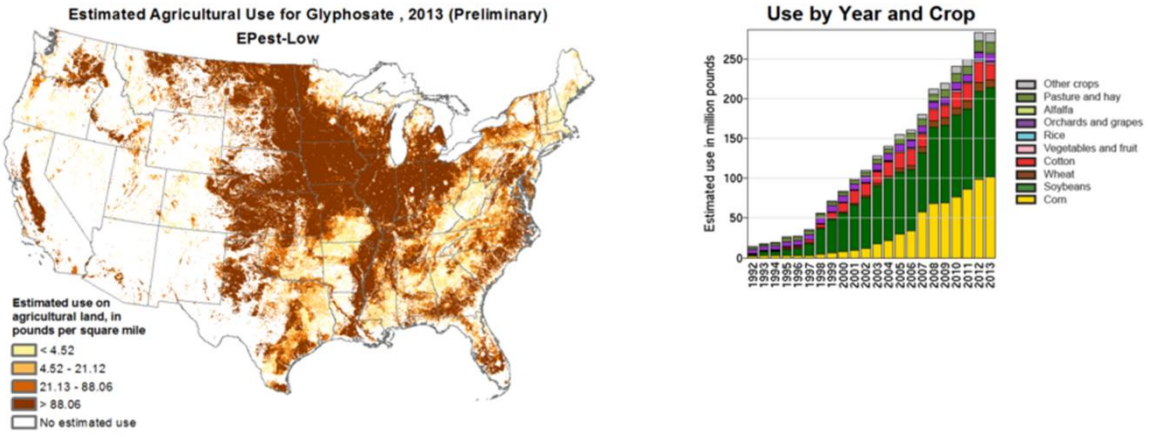


Figure 159. The use of glyphosate (Roundup) in the USA

https://commons.wikimedia.org/wiki/File:Glyphosate_USA_2013.png
Unknown authorUnknown author, Public domain, via Wikimedia Commons

Glyphosate, which is the active ingredient in Roundup and other herbicides, kills plants by interfering with the synthesis of aromatic amino acids (Figure 160.) Glyphosate inhibits an enzyme in the synthesis pathway. (This enzyme does not exist in animal; they obtain aromatic amino acids from their diet.)

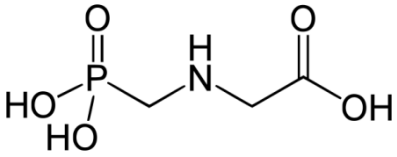


Figure 160. Chemical formula of glyphosate or N-(phosphonomethyl)glycine

Some soil bacteria have a version of this critical enzyme that is resistant to glyphosate inhibition. Scientists have isolated the gene for this enzyme, added plant sequences for transcription, translation and targeting to the chloroplast, and inserted it into several crop plants making them resistant to glyphosate. Roundup Ready plants like e.g. soybean, maize, sugar beets and cotton are widely cultivated in the USA.

Glyphosate resistant plants are only one group among the several herbicide tolerant crop varieties. ***Engineered resistance is now available against a range of herbicides.*** These were the first GM-plants created, and are still the ***most widely grown transgenic crops***.

Insect resistance

Infections by viruses, bacteria and fungi cause severe losses in crop, but probably the most important crop pests are herbivorous insects. Chemical control of insect pests is expensive, and harmful for the environment. Some bacteria produce proteins that kill insect larvae that eat them. On sporulation, the ***bacterium Bacillus thuringiensis*** produces a protein that is poisonous to certain insects. (Spores of this bacterium are widely used in organic gardening.) Genetically engineered crop plants like e.g. maize, cotton, tomato and potato express this ***Bacillus thuringiensis (Bt) toxin*** to ***confer resistance to*** chewing ***insects*** (Figure 161.).

Figure 161. A field of insect resistant maize in Kenya

https://commons.wikimedia.org/wiki/File:Btcornafrica.jpg
Dave Hoisington/CIMMYT, CC BY 2.5 <https://creativecommons.org/licenses/by/2.5>, via Wikimedia Commons

Cry genes coding for the toxin have been isolated from different bacterium strains and extensively modified to suit best expression in plants. The transgenic constructs were supplied with plant promoters, terminators, regulatory elements, polyadenylation signals, and were optimized for plant codon usage. These modified genes have been introduced into plant cells using the Ti plasmid vector. The Bt protein is expressed throughout the plant. On ingestion by an insect, the toxin is activated in the alkaline environment of the gut, and forms holes in the gut wall, eventually killing the insect.

Manipulating male sterility

Hybrid seeds are valuable in agricultural production, because hybrid plants often show increased yield, uniformity and vigor. Hybrid varieties do not breed true; they must be created by crosses between appropriately chosen inbred lines. For many crops, it is either impossible or difficult to generate commercial hybrid seed by conventional means, so scientists turned to genetic engineering for the solution.

Several approaches have been employed for generating transgenic male sterile plants; the first one was the ***barnase/barstar system*** (Figure 162.).

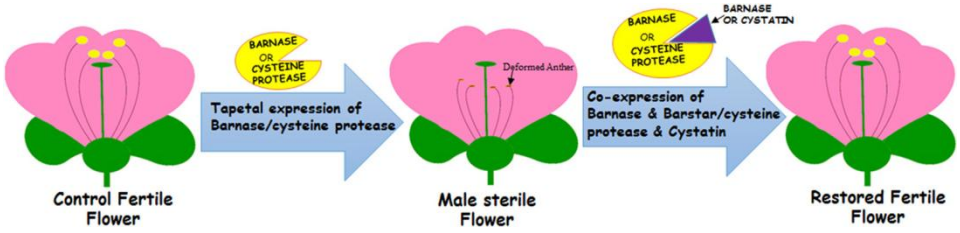


Figure 162. Schematic representation of two male sterility-fertility restoration systems

For a more detailed explanation of the barnase/barstar system, refer to the main text.
Source: Figure 2 in Mol Biotechnol DOI 10.1007/s12033-017-0027-6

The *barnase* gene coding for a ribonuclease from *Bacillus amyloliquefaciens* was fused with a tapetum-specific promoter, and transformed first into oilseed rape and tobacco. ***Expression of the barnase gene kills the cells of the tapetal layer that nourish developing pollen grains***, leading to male sterility. Targeting the expression of *barstar* also to the same cell layer restores fertility in the *barnase*-expressing transgenic plants. When *barnase* expressing male sterile plants are crossed with *barstar*-expressing fertile plants, coexpression of the two genes in the tapetal layer of the anthers of F_1 plants inactivates *barnase* and restores fertility in the F_1 hybrid plant. This technology can also be used to induce sterility in transgenic trees, to prevent gene flow via cross-pollination.

Stress tolerant plants

In traditional agriculture, adjusting the environment to the needs of cultivated crops is a basic endeavor. Farmers try to supply the cultivated plants with sufficient amount of water, improve soil quality with fertilizers, and protect the plants from pests. However, the full potential yield of crops is rarely met, partly because of environmental stresses such as ***drought, cold and salt stress***. Modern biotechnology may help us to shift the relationship between crop plants and the environment, and to modify plants to be able to fend off environmental stresses better.

Scientists apply microarray technologies to identify genes that are up-regulated when plants encounter environmental stresses, including genes involved in stress perception, signaling and tolerance. This approach may reveal the connection between different stress

response pathways, and suggest potential genes for intervention. Studies show, for example, that drought and freezing tolerance may share common protective mechanism.

Salinity is a frequent environmental stress that affects agricultural productivity worldwide, thus several attempts have been made to increase salt tolerance by genetic engineering. One example is the expression of an *Arabidopsis* gene that codes for a sodium/proton antiporter in the transgenic plants. This protein sequesters cytosolic sodium into the vacuole of the plant, thus minimizing sodium toxicity and injury to important enzymes in the cytosol.

Improving nutritional characteristics

Second generation GM plants have been frequently created with the aim of manipulating the quality of the product, that is, improving the nutritional or technological properties of the plant.

Transgenic technologies are applied to several fruits and horticultural plants to ***delay ripening*** and/or ***longer storage life***.

Concerning the nutritional properties of different plants, animal feeds with modified ***amino acid composition***, oil crops producing ***"designer oil"***, or transgenic plants that produce ***modified starch*** for the food processing industry, are all new products of genetic engineering.

For animals and humans, 10 out of the 20 natural amino acids are essential. These amino acids cannot be synthesized by the organisms but must be obtained from the diet. A logical pursuit is, therefore, to attempt to enrich the content of some of these essential amino acids in crop plants. Among the essential ***amino acids, lysine, tryptophan, and methionine*** have received the most attention, because they are the most limiting in cereals (particularly *Lys* and *Trp*), and legumes crops (particularly *Met*). Now several transgenic crops exist that express genes encoding proteins with higher proportion of essential amino acids.

Oil crops are second in importance to cereals as food sources for humans. They provide many industrial products, too. Rapeseed is closely related to the plant model organism *Arabidopsis*, thus it has quickly become the model oil crop. Now a range of transgenic rapeseed varieties exist that produce modified oils (Table XIII). The term "designer oil" has been coined to indicate that different characteristics, such as ***chain length, degree of saturation, or position of double bounds*** have been manipulated in these oils.

Plant	Problem	Solution
Rapeseed	High eruic acid content („antinutritional trait")	Engineering the β-oxidation pathway → LEAR (low eruic acid rapeseed) cultivars → CANOLA
Soybean	Reduce saturated and polyansaturated fatty acid content in favor of monosaturated oils	Manipulating the desaturases of ER → Oleic acid > 85% and saturated fatty acids < 5% of the total fatty acids
Rapeseed	Undesirable fatty acid composition, high amount of C18 fatty acids	Manipulating lipid biosynthesis, introducing 12:0-ACP thioesterase → 40-50% lauric acid content (soap, detergent, surfactants)

Table XIII. Examples for genetic engineering in oil crops

Golden rice: A notable example for ***improving nutritional properties with genetic engineering of the metabolic pathway*** is the golden rice. Golden rice is a variety of rice that is able to synthesize ***β-carotene***, the precursor of vitamin A. This component confers the rice yellow color, hence the name (Figure 163.). Golden rice has been created with the intention to alleviate the vitamin A deficiency of millions of people in the developing world.

Figure 163. Common rice and golden rice

https://commons.wikimedia.org/wiki/File:Golden_Rice.jpg
International Rice Research Institute (IRRI), CC BY 2.0 <https://creativecommons.org/licenses/by/2.0>, via Wikimedia Commons

To produce β-carotene, two transgenes have been integrated into the carotenoid biosynthesis pathway: one from daffodil (*Narcissus pseudonarcissus*), and the other one from the soil bacterium *Erwinia* (Figure 164.). The enzymes coded by these genes are expressed in the endosperm of golden rice. About 300 grams of golden rice a day can supply the β-carotene a person needs; the body can convert it to retinal and subsequently retinol (vitamin A).

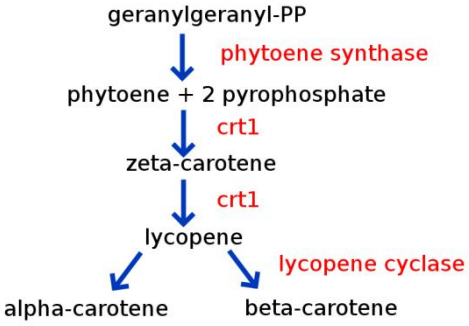


Figure 164. The modified biosynthesis pathway

The gene coding for phytoene synthase (*psy*) comes from daffodil and *crt1* that codes for phytoene desaturase from *Erwinia uredovora*. Lycopene cyclase is produced by the wild-type rice.
https://commons.wikimedia.org/wiki/File:Carotenoidsynthesis.svg
File:Carotenoid.jpg: created by en:user:Petaholmes, modified by Benoit.galletderivative work Smartse, CC BY-SA 3.0 <http://creativecommons.org/licenses/by-sa/3.0/>, via Wikimedia Commons

Plant polymers and biodegradable plastics

Plastics are among the most frequently used materials in everyday life. Most of them are durable and decompose very slowly, causing a more and more severe environmental problem. The production of biodegradable plastics that can be decomposed by bacteria and/or fungi to carbon dioxide and water could be a great benefit to our environment.

The first step towards the realization of this goal was the examination of ***bacteria*** that ***synthesize and accumulate granules of biodegradable plastics called polyhidroxyalkanoates***

(PHA). One such PHA produced commercially is called Biopol, a copolymer of two substances that combines biodegradability with water resistance. The production costs of Biopol and other bacterial PHAs were initially too high to achieve large-scale use in consumer products.

Polyhydroxybutyrate (PHB) is a type of PHAs, synthesized from acetyl-CoA by the sequential action of three enzymes in the bacterium *Alcaligenes eutrophus*. As a next step, scientists investigated the expression of these genes in *Arabidopsis*. Genes for the modified bacterial enzymes were expressed in the plant under the control of a constitutive promoter. Transgenic *Arabidopsis* plants produced and accumulated high amount of PHB in plastids without detrimental effects on plant growth and viability.

Nowadays high yielding ***biomass crops and oil seed crops*** are metabolically engineered for the ***co-production of PHB with lignocellulosic biomass or seed oil***, respectively. Production of PHB in plants provides a renewable cheap source of polymeric material that can be used for the production of plastics, chemicals, and feed supplements. The co-products (lignocellulosic biomass and seed oil) can be utilized to produce energy.

Edible vaccines produced by transgenic plants

Vaccines that improve immunity to particular diseases are valuable innovations of the 19th century research. Traditionally, killed or attenuated strains of the pathogenic organisms have been injected or delivered orally to invoke an immune response.

Second generation vaccines contain only an ***immunogenic protein of the pathogen or subunits of that protein***, which are produced in transgenic microbes. This type of vaccines can also be produced by plants that are genetically modified. The genes encoding the bacterial or viral disease-causing agents are incorporated in plants that have a permanent capacity to express the vaccine. Consumption of these ***edible vaccines*** is then activates the systemic and mucosal immune response against the pathogen. Currently edible vaccines are developed for ***human and veterinary use***. Some examples are summarized in Table XIV.

Plant	Disease/Pathogen	Plant	Disease/Pathogen
Potato	Tetanus, diphteria, hepatitis B, Norwalk virus, enteritis, rabbit haemorrhagic virus	Tobacco	Model plant for developing edible vaccines. Norwalk virus, chicken infectious anaemia, hepatitis B, coccidiosis
Rice	Hepatitis B	Alfalfa	Used to develop edible vaccines mainly for veterinary purposes. Hog pest virus
Banana	Hepatitis B	Carrots	HIV, *E. coli, Helicobacter pylori*
Tomato	SARS, Norwalk virus, *Vibrio cholera* B toxin, hepatitis B, beta-amyloid protein, pneumonia, septicaemia, bubonic plagues	Algae	HPV, HBV
Lettuce	Enteric diseases caused by *E.coli* both in humans and animals		

Table XIV. Edible vaccines

Examples for genetically engineered plants that produce edible vaccines. Edible vaccines are usually in different stages of clinical trials.

Genetic engineering in animals

Transgenic technologies are being employed with many animals, reaching the most advanced results with those ones that are investigated as genetic model systems. The nematode *Caenorhabditis elegans*, the fruit fly *Drosophila melanogaster*, and the mouse *Mus musculus* are the three best-studied animal models used for basic genetic research.

GM livestock for agricultural purposes are being modified with the intention of ***improving economically important traits***, such as growth rate, the quality of meat, milk composition, disease resistance or survival.

Genetically engineering animals is a slow, complicated, and expensive process. Due to the complexity of this problem, the possible ways of introducing foreign DNA into animal cells will be discussed only in a rough copy. Mice will be used as an example, because this species is the most important model for mammalian genetics, and also the best model for investigating human diseases.

Transgenesis in mice: Two strategies for creating transgenic mice are employed: (1) ***ectopic insertion*** and (2) ***gene targeting***.

1. Simple ***injection of the DNA solution into the nucleus of a fertilized egg*** results in ***ectopic insertion*** of the transgene. When injected eggs are inserted into the oviduct of female animals, some will develop into baby mice, having cells that carry the transgene. The transgene becomes integrated into the chromosomes of random nuclei. Typically, an array of the transgene is integrated, but the location, size and structure of the array is different for each integration event. If the integration occurs in germ line cells, the transgene can also be passed on to progeny. Although the expression pattern of the randomly inserted transgene is not always normal and DNA rearrangements can occur inside the arrays, this method is more efficient than gene targeting.

2. ***Gene targeting*** enables researchers to eliminate or modify the function encoded by a gene. Targeted inactivation of a gene is called ***gene knockout***; the application, when a mutant allele is replaced by a wild-type allele (or vice versa) is called ***gene replacement***. These experiments are carried out in cultured embryonic stem cells (ES cells).

To knockout a gene, researchers first construct an inactive, disrupted version of the target gene. The DNA construct containing the defective gene is then injected into the nuclei of cultured ES cells. The defective gene inserts into nonhomologous (ectopic) and homologous sites as well. With a proper selection system it is possible to select those ES cells, where homologous insertion occurred.

In the next step, ES cells that contain one copy of the disrupted gene are injected into early embryos. Adult males develop from these embryos are mated with normal females, and the resulting chimeric progeny are screened for the presence of the modified gene (Figure 165.). Selected individuals are then mated to produce homozygous mice with the knockout in both copy of the gene. The presence of the transgene is then verified using molecular probes for sequences unique to the transgene.

Figure 165.Chimeric parent mouse and two offspring

https://commons.wikimedia.org/wiki/File:ChimericMouseWithPups.jpg
Staff at NIMH's Transgenic Core Facility, Public domain, via Wikimedia Commons

Products from GM animals

Creation of the first GM livestock was reported in 1985, when scientists injected a DNA construct containing the human growth hormone structural gene (hGH) under the mouse metallothionein-I (MT) promoter/regulator region into the nuclei of rabbit, sheep and pig eggs, and verified the integration and expression of the transgene.

Since then, the number of GM farm and fish species has been increasing, serving broadly two purposes: ***they provide different food products or produce specific substances*** in their milk, eggs, blood, etc.

In Table XV, we briefly summarize which animals are most commonly modified, and provide examples of what their genetic modification is for.

Animal	Aim of genetic modification
Swine	Increased growth rate
	Modified milk for better survival and growth of piglets
	Meat with more unsaturated fatty acids
	Cloning GM-pigs with somatic –cell nuclear transfer
	Improved resistance to the influenza virus
	Production of human lysozyme in milk
Goat	Expressing human lysozyme in milk to improve udder health and food safety
	Milk with modified fatty acid content to improve consumer health
Chicken	Increased resistance to diseases (leukosis, bird flue)
	Beta-galactosidase expression for more efficient nutrient utilization and increased Ca-absorption
Cattle	Cloning GM cattle expressing human lysozyme in milk with somatic-cell nuclear transfer
	Increased milk quality: higher level of omega-3 and lower level of omega-6 fatty acids, increased kazein level
	Lysostaphin production in milk: protection from Staphylococcus aureus (mastitis)
	Protection from BSE (no normal cellular prion protein)
Fishes	Growth enhancement (salmonids, tilapia)
	Development of resistance to contagious diseases
	Antifreeze proteins to expand the range of environment suitable for breeding

Table XV. Examples of GM livestock

Pharmaceutical industry and medicine: Transgenic microbes are widely used in the pharmaceutical industry to produce drugs, but they are suitable only to produce relatively simple proteins. Human proteins expressed in mammals are more likely to be similar to their natural counterparts than those expressed in microorganisms. ***Transgenic dairy animals*** (sheep, goats, cattle, pigs, rabbits) are able to produce large amounts of medicinal proteins. They carry gene constructs with a mammary gland-specific promoter, so the recombinant protein is secreted into milk, from which it is easy to purify.

A new field for the application of GM animals is ***xenotransplantation***. Scientists have been experimenting with the creation of genetically modified pigs that can serve as organ donors for humans.

Somatic cell nuclear transfer

Somatic cell nuclear transfer (SCNT) is a method for cloning, when the nucleus of a somatic cell is transferred to the cytoplasm of an enucleated egg (Figure 166.). Then, the ovum that contains the somatic cell's nucleus is stimulated to divide. After many mitotic divisions, the single cell forms a ***blastocyst***, an early stage embryo, whose ***cells contain identical genome to the original nucleus-donating organism.***

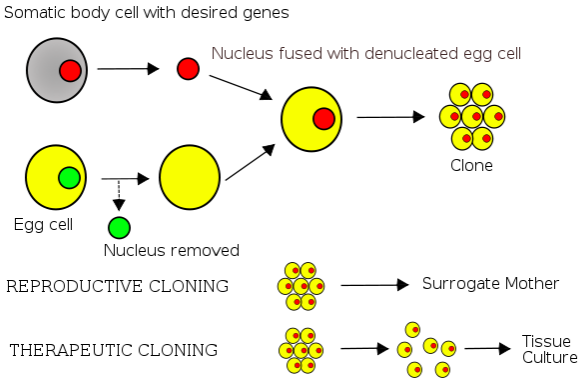


Figure 166. Cloning by somatic cell nuclear transfer

https://commons.wikimedia.org/wiki/File:Cloning_diagram_english.svg
en: converted to SVG by Belkorin, modified and translated by Wikibob, CC BY-SA 3.0 <http://creativecommons.org/licenses/by-sa/3.0/>, via Wikimedia Commons

Reaching the blastocyst stage, researchers have two options to continue the experiment. A part of the blastocyst called inner cell mass (ICM) is a source of ***embryonic stem cells***. By the destruction of the embryo, stem cells can be obtained and cultivated in tissue culture for ***therapeutic*** purposes. If the researchers' goal is ***reproductive*** cloning, the developing embryo is implanted into a surrogate mother and allowed to develop completely.

Dolly: "Dolly" the sheep was the first mammal that was produced by SCNT. The somatic nucleus used to create Dolly derived from a mammary gland (Figure 167.). "She" lived 6 years, and produced several lambs.

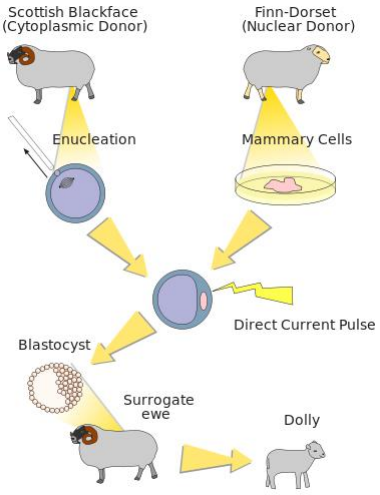


Figure 167. The cloning process that produced Dolly

https://commons.wikimedia.org/wiki/File:Dolly_clone.svg
Squidonius, Public domain, via Wikimedia Commons

In a strict sense, Dolly was not a transgenic animal, as "she" did not carry any genes that was modified or derived from another species.

Somatic cell nuclear transfer can be combined with transgenic technologies and used for several purposes:

1. Genetically engineered large mammals can serve as ***models for human diseases***. Pigs are especially useful models, as they show similarities to humans with respect to organ anatomy, size and physiology.
2. The demand for organ transplants is permanently growing; a possible solution to this problem could be ***xenotransplantation***. Genetically multi-modified pigs are being created, tested, and considered as potential organ donors for humans.
3. In agriculture, the application of transgenic technologies in combination with SCNT could ***improve the health status of livestock***, provide ***healthier and nutritious animal products***, and ***reduce the environmental impact*** of animal production.
4. Scientists have been considering how SCNT technologies could be used in the ***preservation of endangered species.***

Recommended readings

Griffith, Wessler, Lewontin, Gelbart, Suzuki, Miller: Introduction to Genetic Analysis, Eighth Edition. W. H. Freeman and Company, New York, 2005.

Thomas D, Pollard, William C. Earnshow: Cell Biology. Elsevier, Inc. 2004.

Molecular Biology and Biotechnology, Fourth Edition. Ed: John M. Walker and Ralph Rapley. The Royal Society of Chemistry, 2002

Michael R. Davey and Paul Anthony: Plant Cell Culture. Essential Methods. Wiley-Blackwell, 2010

ISBN 978-615-6032-56-0